Article

The Predicted 3D Structure of Human DP Prostaglandin G Protein-Coupled Receptor Bound to CPI Antagonist

Vishnu Shankar, William A. Goddard, Soo-Kyung Kim, and Ravinder Abrol

J. Chem. Theory Comput., **Just Accepted Manuscript** • DOI: 10.1021/acs.jctc.7b00842 • Publication Date (Web): 21 Dec 2017

The 3D Structure of Human DP Prostaglandin G Protein-Coupled Receptor Bound to CPI Antagonist, predicted using the DuplexBiHelix modification of the GEnSeMBLE method

AUTHOR NAMES: Vishnu Shankar, William A. Goddard III[1], Soo-Kyung Kim, Ravinder Abrol, Fan Liu*

AUTHOR ADDRESS**:** [1]Materials and Process Simulation Center (139-74), California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125, USA.

KEYWORDS: protein structure prediction, ligand-protein structure prediction, DP, PGD2, cyclopentanoindole

ABSTRACT: Prostaglandins play a critical physiological role in both cardiovascular and immune systems, acting through its interactions with 9 prostanoid G protein-coupled receptors (GPCRs). These receptors are important therapeutic targets for a variety of diseases including arthritis, allergies, type 2 diabetes, and cancer. The DP prostaglandin receptor is of interest because it has unique structural and physiological properties. Most notably, DP does not have the 3-6 ionic lock common to Class A GPCRs. However, the lack of x-ray structures for any of the 9 prostaglandin GPCRs hampers the application of structure-based drug design methods to develop more selective and active medications to specific receptors. We predict here 3D structures for the DP prostaglandin GPCR, based on the GEnSeMBLE complete sampling with hierarchical scoring (CS-HS) methodology. This involves evaluating the energy of 13 trillion

packings to finally select the best 20 that are stable enough to be relevant for binding to antagonists, agonists, and modulators. To validate the predicted structures, we predict the binding site for the Merck cyclopentanoindole (CPI) selective antagonist docked to DP. We find that the CPI binds vertically in the 1-2-7 binding pocket, interacting favorably with residues R310$^{7.40}$ and K76$^{2.54}$ with additional interactions with S313$^{7.43}$, S316$^{7.46}$, S19$^{1.35}$, etc. This binding site differs significantly from that of antagonists to known Class A GPCRs where the ligand binds in the 3-4-5-6 region. We find that the predicted binding site leads to reasonable agreement with experimental Structure Activity Relationships (SAR). We suggest additional mutation experiments including K76$^{2.54}$, E129$^{3.49}$, L123$^{3.43}$, M270$^{6.40}$, F274$^{6.44}$ to further validate the structure, function, and activation mechanism of receptors in the prostaglandin family. Our structures and binding sites are largely consistent and improve upon the predictions by Li et. al[15] that used our earlier MembStruk prediction methodology.

I. Introduction

Prostaglandins are lipid-based compounds containing 20 carbons, a cyclopentane ring, and several oxy functionalities. Figure 1 shows the prostaglandin D_2 compound (PGD_2) that binds to the prostaglandin DP1 and DP2 receptors. Prostaglandins are associated with a broad array of diseases including asthma, inflammation, cancer, cardiovascular disease, allergies and hypertension[1]. In humans, it is known that presence of an allergen can cause mast cell production of PGD_2. The DP1 receptor has been shown to mediate allergic inflammation, increasing blood flow through vasodilation, thereby contributing to nasal congestion[2]. Additionally, multiple studies[3,4] have verified PGD_2 activity in the brain including sleep regulation, neuroprotection, and hyperalgesia.

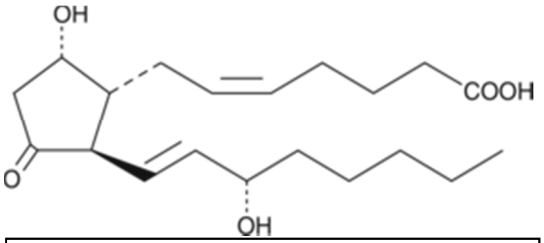


Figure 1. Prostaglandin D_2 Compound (PGD_2)

The prostaglandins function through interactions with one or more of the nine prostaglandin receptors[5], which are G protein-coupled receptors (GPCRs), with seven helical transmembrane domains (TMD). There are eight well-known prostaglandin receptors (DP, EP1-4, FP, IP, and TP), plus a ninth, the recently identified chemoattractant receptor expressed on Th2 cells (CRTH2). The DP receptor is activated by PGD_2, the EP 1, 2, 3, 4, receptors are activated by the corresponding endogenous ligand PGE_2, the FP receptor is activated by PGF_2, the IP receptor is activated by PGI_2, and the other two receptors are activated by thromboxane TXA_2. Unfortunately, there are no x-ray structures for any prostaglandin GPCR, and they are quite distinct from available x-ray structures (Figure 2A) making it difficult to use structural-based approaches to design more effective and selective ligands for therapeutic purposes.

We focus here on the DP1 receptor, which differs substantially from the other prostanoids and is the least understood[6] of the prostaglandin receptors. Accurate 3D structures[60] for these prostanoids receptors would provide the basis for structure-based drug discovery and application, particularly for the development of selective agonists and antagonists that might decrease side effects.

II. Prostaglandin DP Receptor Unique Structural Characteristics

The DP receptor is structurally diverse from all other Class A GPCRs including adrenergic, dopamine, rhodopsin, and adenosine receptors (Figure 2A, 2B)[7]. For example:

- The arginine in the D/ERY motif is conserved in 97%[8] of Class-A GPCRs, but for DP this aligns with the ECW motif.

- The tryptophan (W4.50) on TM4 conserved in Class A GPCRs aligns with serine for the DP receptor.
- The WxP motif on TM6 for Class A GPCRs aligns with SxP on the DP receptor.
- The NPxxY motif on TM7, conserved on Class A GPCRs, corresponds to the DPxxF/Y motif on the DP receptor.

III. Methods

Our strategy for determining the structure of GPCR proteins is to start with a template for the 7 TMDs (based usually on x-ray structures), and then to sample a large number of conformations generated by simultaneous rotations about the axes and tilts of the helix axes. This procedure is called the *GPCR ensemble of structures in membrane bilayer environment* or GEnSeMBLE[11]. As described below, we evaluate the energy of 13 trillion conformations, which we consider a complete set of helix rotations and tilts, but the energy scoring is approximate, being based on adding the pair wise interactions (BiHelix) between helices. Using this approximate energy scoring of all 13 trillion, we reduce to a smaller set of 2000 that are built into 7-helix bundles and rescored to select a final set stable enough to play a role in function. We refer to this type of method as Complete Sampling-Hierarchical Scoring (CS-HS). It involves the following sequence of steps:

- a. PredicTM. Use the alignment of the DP sequence with other GPCRs to predict the hydrophobicity along the sequence which is used to define each TMD and determines the hydrophobic center (HPC) which is taken at z=0 for placement on the x-y mid-membrane plane. Use secondary structure servers to predict extensions of the helical domain beyond the membrane.

- b. Select a template (from an x-ray study or previous predictions) to specify the initial x and y positions on the z=0 plane and the initial tilts (θ, ϕ) with respect to the z axis and prediction of the shape (non-helical distortions) of the TMDs to be used in steps c and d.
- c. Exhaustive (12^7 = 35 million) simultaneous sampling of the rotation angles η (BiHelix) of the seven TMs about their axes.
- d. Exhaustive ($(5*5*3)^7$ = 13 trillion) simultaneous sampling of rotations and tilts: η, θ, ϕ (SuperBiHelix) of the seven TMs.

This procedure has been successful for predicting structures and binding sites for several GPCRs (CCR5[61], CB1[62], AA3[11], Kappa opoid[63], TAS2R38[7], GLP-1R[64], TasTR2/TR3 heterodimers[65]).

A. Prediction of the Initial Helical Bundle

From the PredicTM method[11], we generated a hydrophobicity profile (Figure S2) to predict which part of the DP sequence corresponds to each transmembrane helix (Figure S1). Since the x-ray structures for some GPCRs show clearly that the helical conformations can extend outside the membrane surface[34,41,42], PredicTM also uses secondary structure predictions to predict TMD for which the alpha helices should extend past the membrane. For a more detailed description of how the hydrophobicity profile was generated, please refer to SI Section I.A.

B. Template for the 7-helix bundle and Shape of TMD

B.1 The template. Figure 2A lists the sequence identities to DP for various published structures (both x-ray and predicted). The GPCR with known x-ray structure closest to DP is Dopamine D3 with 15.32% sequence identity (21.03% for TMD). We denote this template as the DRD3 template or DRD3/T.

In addition, we use as a template the structure predicted in 2007[15] using our earlier MembStruk method, which we denote as the DP-Li template or DP-Li/T. This previous MembStruk method

rotated each helix sequentially (after a minimization). We expected that our current method that rotates all helices simultaneously might provide a better structure[7, 10-13, 53, 61-65]. Since the shape was optimized for the same sequence, it is likely to be more accurate than OptHelix.

B.2 The Helix Shapes. Although the dominant structure of the TMD is □lpha-helical, the kinks and distortions distort the shape from a perfect helix. We considered two ways to specify helical shapes:

- For DP-Li/T, we use the shape from the previously optimized structure and the shape obtained from OptHelix MD and minimization denoted as DP-Li/Topt.
- For DRD3/T, we consider the shape from homology denoted at DRD3/T.

The **OptHelix** (55) process starts with the □lpha-helical shape and uses MD to determine the initial structure, which might have kinks induced by helix breakers such as prolines. Specifically, to avoid extraneous distortions due to large polar side chains, OptHelix replaces all residues except Pro, Gly, Ser and Thr with Ala. Then, the TMD is subjected to 2 ns molecular dynamics (MD), after which the average structure is selected, the correct residues restored using SCREAM[14], and the system is energy minimized.

C. BiHelix

Since there is no x-ray structure for any GPCR close to DP (Figure 2A), we consider that the seven rotation angles η might each have any value from 0 to 360º. Previous GEnSeMBLE studies have found that it is sufficient to sample each TMD rotation angles in 30º increments (12 total values). Since we want to sample all possible combinations of rotations, we consider all $(12)^7$ = 35 million conformations. In order to reduce the cost of calculating the energies for 35 million 7-helix bundles, we developed the BiHelix mean field protocol in which the 12 pairs of adjacent helices (1-2, 2-4, 4-5, 5-6, 6-7, 7-1 and all six with coupled with TM3) are examined for

all $(12)^2$ = 144 combinations, where for each combination the side chains are optimized with SCREAM[14] and the structure minimized for 10 steps to eliminate bad contacts. We then use these 12*144=1728 pair-wise energies, to estimate the energy for all 35 million combinations, ignoring any incompatibilities of the side chains from one pair with another (mean field). We then select the lowest 2000 for constructing the full 7-helix bundles, again optimizing the side chains using SCREAM[14] followed by 10 steps of minimization, a procedure referred to as ComBiHelix. From this set of 2000, we select the lowest 10 to 20 for further analysis. We have found that applying this approach to a known x-ray structure essentially always finds the x-ray structure[10] to be the lowest in ComBiHelix, validating the use of the scoring function (Dreiding Energy with SCREAMed side chains) and the BiHelix procedure to identify the best packings.

No.	Structure Compared	All	TM Avg.	TM1	TM2	TM3	TM4	TM5	TM6	TM7
1	PD2R_HUMAN	100	100	100	100	100	100	100	100	100
2	DRD3_HUMAN	15.32	21.03	20.83	38.46	28.57	10.53	25.00	23.81	0.00
3	AA2AR_HUMAN	13.93	20.28	8.33	19.23	38.10	21.05	16.67	28.57	10.00
4	5HT2B_HUMAN	13.65	18.50	12.50	26.92	33.33	15.79	16.67	14.29	10.00
5	ADRB1_MELGA	13.37	18.18	25.00	19.23	14.29	31.58	8.33	23.81	5.00
6	S1PR1_HUMAN	13.09	17.58	8.33	19.23	23.81	31.58	20.83	14.29	5.00
7	ADRB2_HUMAN	12.53	16.79	16.67	19.23	14.29	21.05	12.50	23.81	10.00
8	CNR1_HUMAN	11.98	17.26	29.63	23.33	10.34	12.00	12.50	15.62	17.39
9	AA3R_HUMAN	11.98	18.01	12.50	15.38	28.57	15.79	25.00	23.81	5.00
10	CXCR4_HUMAN	11.98	16.08	12.50	23.08	28.57	21.05	8.33	19.05	0.00
11	HRH1_HUMAN	11.70	8.69	8.33	19.23	9.52	5.26	4.17	14.29	0.00
12	OPSD_BOVIN	11.42	18.24	12.50	19.23	23.81	26.32	12.50	33.33	0.00
13	M2MR_HUMAN	10.86	11.40	20.83	11.54	19.05	10.53	8.33	9.52	0.00
14	ACM3_RAT	10.58	12.71	20.83	11.54	19.05	10.53	12.50	9.52	5.00
15	OPRX_HUMAN	10.31	13.99	16.67	15.38	23.81	10.53	12.50	19.05	0.00
16	GLP1R_HUMAN	10.31	10.45	0.00	11.54	19.05	10.53	12.50	9.52	10.00
17	OPRD_MOUSE	10.03	11.96	12.50	15.38	19.05	0.00	12.50	14.29	10.00
18	OPRM_MOUSE	9.47	10.92	12.50	15.38	14.29	15.79	4.17	14.29	0.00
19	CCR5_HUMAN	9.47	7.94	12.50	11.54	14.29	0.00	12.50	4.76	0.00
20	OPRK_HUMAN	8.91	8.27	8.33	11.54	14.29	5.26	4.17	14.29	0.00
21	MRGX1_MOUSE	8.64	11.52	16.67	15.38	14.29	10.53	0.00	23.81	0.00

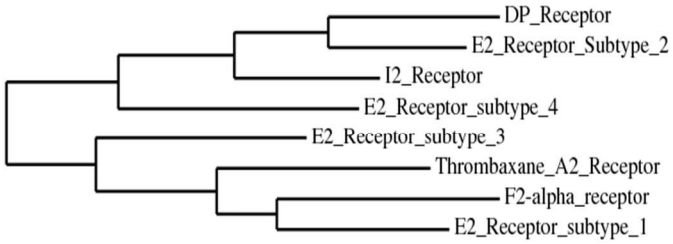


Figure 2. A. (Left) The sequence identity of the prostaglandin DP receptor with other G protein-coupled receptors (GPCRs) for which there are X-ray structures or with accurately predicted structures. The highest homology is with dopamine but it is only 15%. **B**. (Right) The phylogenetic tree with the other prostaglandin receptors shows that DP receptors are the most related to PE2R2, but DP is one of the least similar and studied receptors of the prostanoid group.

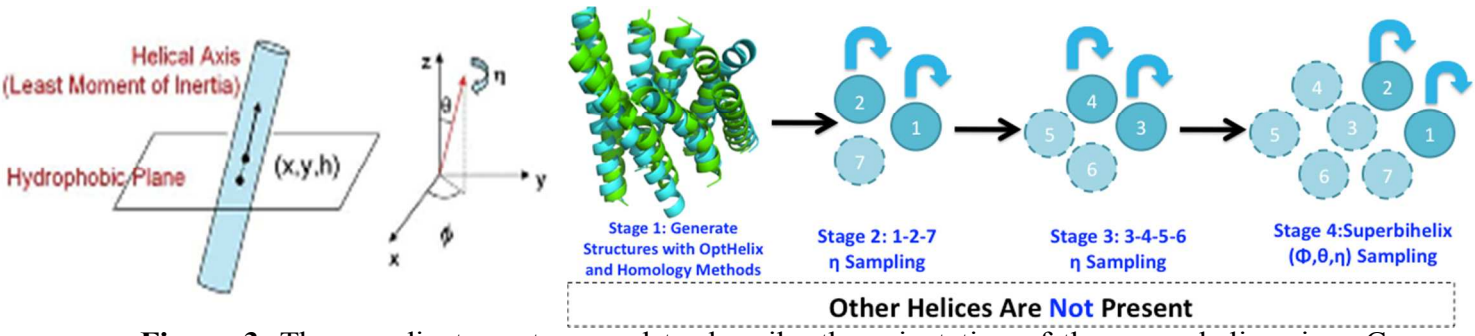


Figure 3. The coordinate system used to describe the orientation of the seven helices in a G protein-coupled receptor (GPCR) bundle. Taking a common x-y plane for the hydrophobic center of each helix, θ is the tilt of the helix axis with respect to the z axis while φ is the projection of this axis on the x-y plane (where φ=0 is the x axis); and η is the rotation about the helix axis. The x and y are the coordinates of the helix axis intersecting the plane, where (x,y) = (0,0) is the projection of the central axis of TM3 onto the xy plane and the y axis points to the projection of TM4 onto this plane. The BiHelix analysis examines the interactions between the

12 strongly interacting helices. Double arrows connecting nearest neighbor helix pairs that are sampled independently in the BiHelix[10] procedure. The standard BiHelix procedure was modified for the DP structural analysis. Rather than sampling the rotations of all 7 helices simultaneously, we first sampled all helix rotations of the 1-2-7 helices simultaneously, and then selected 4 or 5 diverse cases from the 10 lowest energetic structures for sampling the 3-4-5-6 helix rotations simultaneously. The reason for this change in the standard BiHelix procedure is that the protonated water (or Na^+) likely to compensate for the charge on the $D72^{2.50}$ in the middle of TM2 is ignored in the BiHelix analysis, making recognition of the most favorable 1-2-7 packings more difficult. This dual sampling process DuplexBiHelix allows the best 1-2-7 packings to accommodate this H_3O^+ to be more clearly recognized.

In comparing energies, we consider two ways to evaluate energies:

- **Interhelical**: The sum of the 12 sets of interhelical interactions (ignoring intrahelical contributions)
- **Total energy**: The sum of all energetic interactions (including interhelical and intrahelical contributions)

We find that intrahelical energies are sensitive to the helix shape obscuring the choice for the best packing. Thus, we consider that the interhelical energies are more sensitive for selecting the optimal packing of the TMD. Even so, we also examine the cases with best total energy.

For each of these two sets of energy we consider two models for evaluating the energies

- **Charge model** (Normal Coulomb): each Lys and Arg residue have a net charge of +1 while each Asp or Glu have a net charge of -1 and each unprotonated His has a charge of 0.
- **Neutral model**: For each salt bridge the SCREAM[14] program transfers the proton from the acceptor back to the donor so that the residues are neutral. For isolated charged residues, the residues are still neutralized by adding or subtracting a proton.

With charged residues, the long-range of the Coulombic interactions can sometimes result in major re-ordering in packing energies for relatively unimportant changes in side chain placements of remote residues[56]. Such biases are removed in the neutralized model. We generally rank the structures using all four criteria, but with more emphasis on the neutral energies, which has been demonstrated to select the best final structure for many predicted systems[7, 11-13].

An important issue for many class A GPCRs is that the conserved $Asp^{2.50}$ in the middle of TM2, usually has no nearby positive residue with which to form a salt bridge. Thus, we expect

that the real structure will have a charge compensating H_3O^+ or Na^+ coupling to the conserved Asp and perhaps also the Asn in TM1 and TM7 that is often coupled to the Asp[43-45]. However, our scoring criteria for BiHelix does not include this positive center, which we expect may cause noise in estimates for the energy orderings. Even so most of our previous predictions of lowest energy GPCR structures identify correctly the 1-2-7 coupling in most of the low energy conformations. For DP there is an additional complication because the positive $Lys^{2.54}$ is just one turn up from the conserved $Asp^{2.50}$ residue. Consequently, we find that the lowest 2000 structures from BiHelix leads to very few cases with strong 1-2-7 couplings or even 2-7 interactions. In order to better identify 7 helix bundles that have strong 1-2-7 couplings, we modified the standard BiHelix procedure by breaking it into 2 steps as follows (Figure 3):

- First, we carried out BiHelix for the $(24)^3$=13,824 combinations of the 1-2-7 helices (using 15º increments in η), from which we selected the best 1000 for CombiHelix.
- Then, for each of these top 1000 structures, we used BiHelix to estimate the energies for the $(24)^4$ ~ 330,000 combinations of TMD 3, 4, 5, 6. Then, the energies of these ~4 billion 7-helix systems were compared to select the best 2000.
- Finally, CombiHelix was used to evaluate the energy for these 2000 7-helix bundles, from which we selected the 20 best for docking.

This DuplexBiHelix procedure finds many more configurations in the best 20 that have good interactions among the 1-2-7 helices. For instance, we find starting with DP-Li/T led to 9/20 structures with strong 1-2 interactions $N34^{1.50}$ – $K76^{2.54}$ and 11/20 structures with 2-7 interactions, namely either $D72^{2.50}$ – $S316^{7.46}$ or $K76^{2.54}$ – $S316^{7.46}$. Additionally, we find 17/20 structures, where the conserved $D72^{2.50}$ forms a salt-bridge with $K76^{2.54}$.

We carried out this DuplexBiHelix procedure using all three templates: DP-Li/T, DRD3/T, and DP-Li/Topt leading to 2000 structures from which we picked the best 20 CombiHelix structures listed in SI Table 1 and 2. (SI Table 1 and 2 lists the lowest energy structures for each of the selected templates).

After sampling TM1, 2, or 7, we see the initial minimized DP-Li/T and DRD3/T templates with 0 rotation for TM1, 2, and 7 is in the best 10 structures. For DP-Li/Topt, we see the top 10 structures are mixed in that some of the structures have large changes (255° – 285°) in at-least one of helices for TM1, 2, or 7 compared to the template, whereas other structures have smaller changes (15° - 45°) compared to the template structure (SI Table 1). We note that for the DRD3/T and DP-Li/Topt top structures with smaller changes in TM1, 2, or 7 compared to the template, the structures form an identical 1-2-7 interaction, when compared to the DP-Li/T case. For DP-Li/Topt structures with larger changes in TM1, 2, or 7, these structures form an identical $D72^{2.50}$-$K76^{2.54}$ salt bridge, but have different 1-2-7 interactions compared to DP-Li/T, including $T29^{1.45}$ – $K76^{2.54}$ and $K76^{2.54}$ -$S313^{7.49}$.

In analyzing the top structures on all 3 templates DP-Li/T, DRD3/T, and DP-Li/Topt after our DuplexBiHelix procedure, we note some common interactions. In addition to the shared residues for the 1-2-7 interhelical interactions ($N34^{1.50}$, $D72^{2.50}$, $K76^{2.54}$, $S316^{7.46}$), we find that the top structures from all 3 templates share the following favorable salt-bridges: $D72^{2.50}$ - $K76^{2.54}$, $R284^{6.54}$ – $E304^{7.34}$, $K291^{6.61}$ – $D292^{6.62}$. We find the $D72^{2.50}$ - $K76^{2.54}$ salt bridge is of significance, since our DuplexBiHelix procedure aims to find a positive charge partner for the Asp in the middle of TM2 and since $K76^{2.54}$ is unique to the DP receptor. Additionally, as our alignment indicates (Figure S1), both $R284^{6.54}$ – $E304^{7.34}$ and $K291^{6.61}$ – $D292^{6.62}$ are not

conserved in the prostaglandin receptor family, which indicates that these salt-bridge interactions uniquely provide functional stability to TM6 and TM7 of the DP receptor.

Unique to the DP-Li/Topt top-ranked structures, we find a large rotation about TM3, 4, and 5 (SI Table 2). Of note, we find largest changes for TM3 in DP-Li/Topt compared to the other two templates. As a result, we find a unique strong (-42.87 kcal/mol for the top-ranked DP-Li/Topt receptor) salt-bridge interaction between E129$^{3.49}$ – R223$^{5.55}$, where R223$^{5.55}$ is unique to the DP receptor. In comparing DP-Li/Topt and DP-Li/T, we find this TM3-5 interaction comes at the cost of other TM3-7 interactions found in the DP-Li/T top structures such as D319$^{7.49}$– S119$^{3.39}$/Q122$^{3.42}$, which were previously predicted in agonist and antagonist dynamic studies to be critical for receptor stability and potentially DP activation[15].

Within the DRD3/T top-ranked structures, we find large rotations of TM5 and relatively minor rotations for TM3, 4, and 6. Due to the large change about TM5, we find the formation of many favorable intrahelical hydrogen bonds on TM5 and new interactions between TM3 and 5. For instance, we find strong intrahelical hydrogen bonds between N218$^{5.50}$ – L219$^{5.51}$, L219$^{5.51}$ – R223$^{5.55}$, and T214$^{5.46}$ – N218$^{5.50}$, which would likely contribute functional stability as the receptor undergoes conformational changes. We also find unique hydrogen bond interactions between TM3 – 5, such as W131$^{3.51}$ – Q233$^{5.65}$, where both the bond acceptor and donor are unique to the DP receptor within prostaglandin family.

Looking at the DP-Li/T top-ranked structures, we see that the top 10 structures have big changes in the rotation about TM4 (SI Table 2) with some changes in TM 3, 5, 6, but all maintain the initial 1-2-7 coupling. This gives confidence in the interhelical contacts and position of residues involved in binding, while providing a better arrangement of the residues on TM4. Within the DP-Li/T top structures compared to the other two templates, we find a unique

favorable interaction (-7.7 kcal/mol) between R230$^{5.62}$ – H263$^{6.33}$, where both charged partners are conserved in the hE2R2 and hIP prostaglandin receptor family. In addition, much like the DRD3/T top-ranked structures, we find strong intrahelical and interhelical interactions on TM5 and TM6, including T214$^{5.46}$ – N218$^{5.50}$, R223$^{5.55}$ – T271$^{6.41}$, T275$^{6.45}$ – M276$^{6.46}$.

In summary, due to our DuplexBiHelix methodology, we are able to find favorable 1-2-7 interactions and unique interhelical interactions with residues on TM3, 4, 5, 6 in each template. However, since DP-Li/T top-ranked structures are able to capture both favorable intrahelical interactions (found in DRD3/T top-ranked structures) and strong interhelical hydrogen bonds, we conclude that the top-ranked DP-Li/T structures are more favorable compared to structures derived from DP-Li/Topt and DRD3/T. In addition to DP structural predictions, the DuplexBiHelix modification of GEnSeMBLE substantially improves the accuracy while reducing the cost. This should be valuable for other applications to GPCRs structure prediction. In particular, sampling the TM1-2-7 coupling first can better account for the charged Asp$^{2.50}$ residue in the middle of TM2, which may actually be coordinated to H_3O^+ or Na^+.

D. SuperBiHelix

The BiHelix procedure optimizes the contacts between the helices for rigid rotations, but we know from comparisons of various x-ray structures that the tilts from the z-axis (θ, relative perpendicular to the membrane) can change by 10° and that the azimuthal angle φ for this tilt in the membrane plane can change by up to 40°[10]. Thus we must also consider the helix tilts (θ, φ) simultaneously with the rotation (η). We do this using the SuperBiHelix procedure in which each of the best combinations of rotations from BiHelix is followed by simultaneous changes in all three angles. Previous studies showed that applying SuperBiHelix to the known x-ray structure

of one GPCR leads correctly to the known x-ray structure for another GPCR as the lowest energy[66].

We applied the SuperBiHelix procedure to the 5 most diverse of the top 10 DuplexBiHelix structures (structures 1, 4, 6, 8, 10 from Table S2) for the DP-Li/T structures. We also carried out SuperHelix on a diverse set of structures from the DuplexBiHelix level for the DP-Li/Topt structures (1, 3, 9, 10) and DRD3/T structures (1, 5, 10). Here we consider simultaneous variations of

- 0, ±10 in θ,
- 0, ±20, ±40 in φ, and
- 0, ±15 in η

leading to $(3*5*3)^7$ = 373 billion combinations.

We evaluated the energies for all 373 billion combinations using the DuplexBiHelix methodology of SuperBiHelix and selected the best 2000. We built these 2000 into 7-helix bundles, used SCREAM[14] to optimize side chains, and carried out limited minimization. Finally, based on the combinations of the rankings for the four sets of energies, we selected the best 20 7-helix bundle conformations shown in Table 1 as the ensemble of apo-protein structures, which we used for ligands.

In summary, since we find that strong 1-2-7 interhelical interactions, stabilizing salt-bridges, 3-7 interhelical, and intrahelical interactions in DP-Li/T structures are maintained, we predict that these ranked structures will produce the most optimal predicted binding mode. In SI Section II.A, we have conducted a more detailed comparison of the DP-Li/T, DRD3/T, DP-Li/Topt structures. To further understand which receptor interactions are most likely in the

inactive receptor, we docked the ligand to a selected subset of receptors from the SuperBiHelix level for all 3 templates.

receptor variants at the SuperbiHelix level	top n	Rank from DuplexBiHelix Procedure	θ							φ							η							CInterH	CTotal	NInterH	NTotal
			H1	H2	H3	H4	H5	H6	H7	H1	H2	H3	H4	H5	H6	H7	H1	H2	H3	H4	H5	H6	H7				
Previous Computational MD DP Model	1	1	10	0	-10	0	-10	0	0	15	-30	0	15	0	15	15	0	-15	0	-15	0	0	0	-425.973	-225.349	-353.296	-203.036
Neutral Total Energy Ranks	2	1	10	0	-10	0	-10	0	0	15	-30	0	15	0	15	15	0	-15	0	0	0	0	0	-429.468	-219.599	-358.229	-196.996
Top Structures (DP-Li/T)	3	1	10	0	-10	-10	-10	0	0	15	-30	0	15	0	15	15	0	-15	0	30	0	0	0	-421.999	-224.559	-346.709	-196.716
	4	1	10	0	-10	0	-10	0	0	15	-30	0	30	0	15	15	0	-15	0	0	0	0	0	-421.536	-208.839	-346.449	-192.080
	5	1	10	0	-10	10	-10	0	0	15	-30	0	30	0	15	15	0	-15	0	0	0	0	0	-421.435	-219.780	-347.394	-191.188
	6	8	0	-10	-10	0	0	0	-10	0	0	30	0	15	-30	-15	30	0	30	15	-30	15	-15	-416.474	-204.899	-334.066	-190.200
	7	1	10	0	-10	10	-10	0	0	15	-30	0	30	0	15	15	0	-15	0	15	0	0	0	-424.462	-214.348	-350.622	-189.836
	8	4	0	-10	-10	10	0	0	-10	0	-15	15	15	0	0	0	0	0	0	0	-15	15	0	-402.400	-190.459	-339.178	-187.321
	9	4	-10	-10	-10	10	0	0	-10	0	30	15	0	0	0	-15	0	-15	15	-15	-15	15	-15	-411.305	-188.299	-356.999	-186.258
	10	6	0	0	0	-10	0	0	-10	0	0	0	-15	0	0	15	0	0	0	15	-15	15	0	-405.315	-208.828	-347.849	-185.656
	11	8	0	-10	-10	0	-10	-10	-10	0	0	15	0	15	-30	-15	15	0	30	0	-15	15	-15	-366.314	-185.158	-298.117	-185.360
	12	8	-10	-10	-10	0	0	0	-10	0	30	30	0	15	-30	-15	30	0	30	0	-30	15	-15	-422.042	-196.169	-332.056	-182.133
	13	8	0	-10	-10	10	0	0	-10	0	-15	30	30	15	-30	-15	30	0	30	0	-30	15	-15	-410.598	-201.089	-318.903	-181.581
	14	6	0	-10	-10	10	0	0	-10	0	0	15	15	0	0	-15	15	0	30	-15	-30	30	-15	-387.316	-177.999	-314.030	-181.062
	15	8	0	-10	-10	0	-10	-10	-10	0	0	15	0	15	-30	-15	15	0	30	15	-15	15	-15	-370.044	-188.870	-303.429	-180.947
	16	8	0	-10	-10	0	0	0	-10	0	0	30	0	15	-30	-15	30	0	30	30	-30	15	-15	-411.968	-198.338	-328.925	-180.825
	17	1	10	0	-10	0	-10	-10	-10	15	-30	-15	15	0	0	15	0	-15	0	0	-15	0	0	-431.208	-203.898	-365.255	-180.570
	18	6	0	-10	-10	0	0	0	-10	0	0	15	0	0	0	-15	15	0	30	-30	-30	30	-15	-380.001	-181.068	-303.736	-180.019
	19	8	0	-10	-10	10	0	0	-10	0	0	30	30	15	-30	-15	15	0	30	0	-30	15	-15	-406.144	-205.860	-321.193	-179.529
	20	1	10	0	-10	0	0	0	0	0	-15	0	0	0	15	15	15	-30	0	-30	-15	0	0	-414.024	-198.070	-346.804	-179.089
GeNSEMBLE DP OptHelix Model	1	10	-10	0	0	0	-10	-10	-10	-15	15	0	0	-15	-15	-15	0	0	0	15	30	30	-30	-455.137	-517.769	-341.227	-502.160
Neutral Total Energy Ranks	2	10	0	0	0	10	-10	-10	-10	0	0	0	15	-15	-30	-30	0	0	0	15	30	30	0	-432.668	-529.230	-330.186	-501.530
Top Structures (DP-Li/Topt)	3	10	-10	0	0	0	-10	-10	-10	-15	15	0	15	-15	-30	15	0	0	0	15	30	30	-30	-418.466	-513.279	-331.354	-501.099
	4	10	-10	0	0	0	-10	-10	-10	0	15	0	15	-15	-15	-30	15	0	0	15	15	30	-15	-483.801	-511.118	-348.120	-498.837
	5	10	0	0	0	0	-10	-10	-10	0	0	0	15	-15	-30	-30	0	0	0	15	30	30	0	-431.149	-521.360	-332.617	-498.319
	6	10	-10	0	0	0	-10	-10	-10	0	15	0	0	-15	-30	-30	15	-15	0	15	30	30	-30	-473.837	-520.480	-356.634	-496.069
	7	10	-10	0	0	10	-10	-10	-10	0	15	0	15	-15	-15	0	15	0	0	15	15	30	-30	-490.023	-521.078	-362.484	-495.435
	8	10	-10	0	0	0	-10	-10	-10	-15	15	0	15	-15	-15	-15	0	0	0	30	30	30	-30	-454.290	-507.599	-341.969	-493.892
	9	10	-10	0	0	0	-10	-10	-10	0	15	0	15	-15	-15	-15	0	0	0	15	15	30	-30	-454.047	-510.110	-346.187	-493.239
	10	10	-10	0	0	0	-10	-10	-10	-15	30	0	0	-15	-30	-30	0	0	0	15	30	30	-30	-480.265	-487.299	-347.841	-493.182
	11	10	-10	0	0	0	-10	-10	-10	-15	15	0	0	-15	-30	15	0	0	0	15	30	30	-30	-426.609	-506.249	-339.136	-492.733
	12	10	-10	0	0	0	-10	-10	-10	0	15	0	30	-15	-15	-15	0	0	0	0	15	30	-30	-458.672	-502.879	-338.908	-492.165
	13	10	-10	0	0	0	-10	-10	-10	-15	15	0	15	-15	-15	15	0	0	0	15	15	30	-30	-454.482	-528.329	-350.717	-491.930
	14	10	-10	0	0	0	-10	-10	-10	0	15	0	30	-15	-30	-30	0	0	0	30	30	30	0	-424.991	-496.540	-318.897	-491.514
	15	10	-10	0	0	0	-10	-10	-10	-15	15	0	0	-15	-15	-15	0	0	0	15	15	30	-30	-470.541	-520.680	-349.157	-490.693
	16	10	-10	0	0	0	-10	-10	-10	-15	15	0	0	-15	-30	-15	0	0	0	15	30	30	0	-465.005	-496.179	-346.144	-489.987
	17	10	-10	0	0	0	-10	-10	-10	0	15	0	15	-15	-30	-30	0	0	0	15	30	30	0	-435.088	-485.810	-334.102	-489.908
	18	10	-10	0	0	0	-10	-10	-10	0	15	0	15	-15	-30	-30	15	-15	0	0	30	30	-30	-469.810	-508.128	-358.969	-489.134
	19	10	-10	0	0	10	-10	-10	-10	-15	15	0	15	-15	-15	15	0	0	15	15	15	0	-30	-430.489	-538.160	-320.658	-489.105
	20	10	-10	0	0	0	-10	-10	-10	0	15	0	0	-15	-30	-30	0	0	0	15	30	30	0	-450.795	-478.310	-347.810	-488.636
Dopamine DRD3 DP Homology Model	1	10	0	0	0	0	0	-10	0	0	0	0	-30	0	-15	0	-15	0	0	-15	0	0	30	-444.434	-348.699	-401.306	-353.459
Neutral Total Energy Ranks	2	10	0	0	0	0	0	-10	0	0	0	0	-30	0	-30	0	-15	0	0	-15	0	0	30	-428.620	-338.349	-382.661	-344.114
Top Structures (DRD3/T)	3	10	0	0	0	0	0	-10	0	-15	0	0	-30	0	30	0	0	15	0	-15	0	15	15	-433.389	-328.459	-384.886	-341.872
	4	10	10	0	0	0	0	-10	10	-15	0	0	-30	0	-15	15	-15	0	0	-15	0	15	30	-430.121	-325.659	-376.854	-339.339
	5	1	10	0	0	0	0	0	10	-15	-15	0	-30	15	-30	15	-15	0	0	0	-15	-15	30	-410.890	-299.790	-358.349	-334.957
	6	10	0	0	0	0	0	-10	0	0	0	0	-30	0	-30	0	0	0	0	-15	0	0	30	-437.022	-324.019	-390.107	-332.928
	7	1	0	0	0	0	0	0	10	-15	0	0	-30	15	-30	15	-30	0	0	0	0	-15	30	-446.642	-330.550	-348.532	-332.785
	8	1	0	0	0	0	0	0	0	0	0	0	-30	-15	0	0	0	0	15	-15	0	0	0	-434.588	-319.318	-373.121	-332.301
	9	10	10	0	0	0	0	-10	10	-15	0	0	-30	0	-30	15	-15	0	0	-15	0	15	30	-433.699	-333.219	-387.640	-330.033
	10	1	10	0	0	0	0	0	10	-30	0	0	-30	15	-30	15	0	0	0	0	0	-15	30	-418.328	-306.150	-329.662	-328.359
	11	1	0	0	0	0	0	0	0	0	0	0	-30	-15	0	0	0	15	0	0	0	0	0	-444.775	-332.429	-382.527	-327.815
	12	10	0	0	0	0	0	-10	10	-30	0	0	-30	0	-15	15	15	0	0	-15	0	0	15	-449.665	-327.229	-372.668	-327.191
	13	1	0	0	0	0	0	0	10	-15	0	0	-30	-15	0	15	-15	0	0	0	0	0	0	-419.554	-310.718	-362.598	-327.020
	14	10	0	0	0	0	0	-10	10	-15	0	0	-30	0	-15	0	0	0	0	-15	0	30	30	-409.786	-322.548	-358.921	-326.915
	15	10	0	0	0	0	0	-10	10	-15	0	0	-30	0	-15	15	-30	0	0	-15	0	15	30	-465.029	-345.749	-384.871	-326.537
	16	1	0	0	0	-10	10	0	10	-15	-15	0	30	15	-30	15	-30	0	0	-15	0	-15	30	-416.538	-293.039	-353.980	-325.894
	17	10	0	0	0	0	0	-10	-10	0	0	0	-30	0	15	-15	-15	0	0	-15	0	0	0	-452.556	-340.989	-380.279	-325.428
	18	10	0	0	0	0	0	-10	10	-15	0	0	-30	0	-15	15	0	0	0	-15	0	15	30	-445.240	-341.739	-373.948	-325.305
	19	1	10	0	0	-10	10	0	10	-15	-15	0	30	15	-30	15	-15	0	0	-15	0	-15	30	-382.798	-273.370	-326.651	-324.674
	20	10	10	0	0	0	0	-10	10	-15	0	0	-30	0	-15	15	-15	-15	0	-15	0	15	30	-418.604	-324.030	-356.672	-323.499

Table 1. SuperComBiHelix Results for DP Receptor with: (a) DP Receptor based on previous computational MD DP structure (DP-Li/T), (b) GEnSeMBLE OptHelix DP helices (based on the previous computational DP template (DP-Li/Topt) (c) Dopamine (DRD3) Homology Template (DRD3/T) φ, θ, and η Values.

E. Docking

Next, we docked the CPI antagonist to a subset of the ensemble of 20 apo proteins from SuperBiHelix (θ, φ, η sampling) for all 3 helical templates (DRD3/T, DP-Li/T, and DP-Li/Topt) based on a subset of structures from the top 20 wild-type (WT) neutral total energy rankings (Table 1). Ultimately, we decided on the structures from SuperBiHelix to use in docking based on two criteria. One, we selected diverse structures based on root-mean squared deviation (RMSD) to maximize the chance of having different binding sites. Two, we considered the inter-helical interactions found in each top-ranked structure, favoring cases with greater number of interhelical hydrogen bonds and salt-bridge contacts.

Then, for each structure selected we used the Darwin Dock Complete Sampling - Hierarchical Scoring (CS-HS) protocol[11,57], which samples iteratively a large number (~50,000) poses without scoring, then collects them into Voronoi families based on RMSD=2Å, energy scores only the family heads, and selects the best 10% of the families for evaluating the energies of their children. This allows a rather complete sampling of the binding site, while minimizing the number of energy calculations. This docking analysis is carried out by for the alanized protein in which the six hydrophobic residues (I, L, V, F, Y, and W) are replaced with Ala. Finally, we select the best 100 ligand poses, for each of which we add back the hydrophobic residues using SCREAM[14]. Thus, each of the 100 poses has its own individually optimized side chains to accommodate the pose. After minimization, we pick the final best pose or two based on energy. Then the ligand-protein complex is neutralized, and the protein-ligand complex undergoes full geometrical minimization.

The above predictions did not include any explicit solvation. We selected the predicted lowest energy antagonist-bound receptor complex to further relax and test the stability of

predicted interactions under physiological conditions. For the selected complex, our receptor is from the SuperCombiHelix top-ranked neutral total energy structures of the DP-Li/T template. We selected this complex, since among the docked SuperCombiHelix neutral total energy top-ranked receptor complexes, it had one of the lowest complex total energies and favorable interactions with the antagonist (discussed in Section IV.E). We used the interactions between the binding site and receptor to compare antagonist-bound receptor complexes across structures from different helical shape templates.

Based on the unique inter-helical and favorable ligand-receptor interactions, we selected the DP-Li/T docked structures for further relaxation in dynamics. In SI Section II.B, we have compared the docked structures for DP-Li/T, DP-Li/Topt, DRD3/T in greater detail and have elaborated on why DP-Li/Topt and DRD3/T structures were not selected for further relaxation.

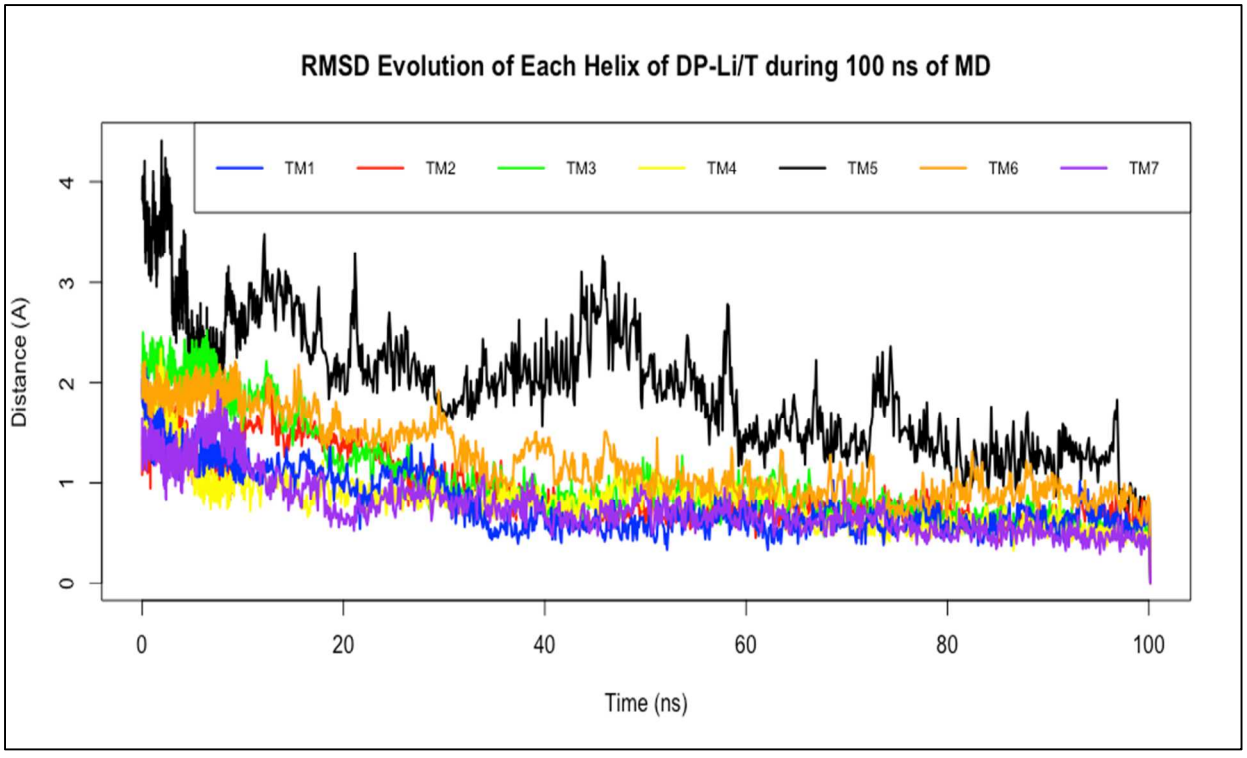


Figure 4. RMSD Evolution of each helix backbone of selected DP-Li/T complex during 100 ns of MD. The last frame of MD was selected as reference frame.

Figure 4 shows the root-mean square deviation (RMSD) change of the helix backbone of selected DP-Li/T complex during 100 ns of MD. The RMSD is with respect to the last frame of the 100 ns trajectory. The total range of RMSD during the 100 ns MD is 0-4.41 Å. In the last 20 ns of MD, the total range of RMSD for the complex backbone is 0-1.87 Å. When focusing on the RMSD changes of each helix during 100 ns of MD, we find that that all transmembrane (TM) regions except TM5 has a total RMSD range from the initial structure from 0 – 2.52 Å. TM5 RMSD fluctuations range from 0 – 4.41 Å, with changes ranging from 0 – 1.83 Å in the last 20 ns of MD. This is to be expected, since TM5 is the longest (35 residues) compared to the other TM regions in the receptor, meaning it likely has greater conformational flexibility. Figure 7 (Results & Discussion) also shows the stability of critical hydrogen bonds during 100 ns of MD. The RMSD evolution data in conjunction with the stability of important hydrogen bonds implies that 100 ns is sufficiently long to reach a stable protein-ligand complex. The MD setup, parameters, protocol are described next:

F. Molecular Dynamics (MD)

F1. Setup of the Protein-Lipid bilayer system for molecular dynamics

The protein-ligand complex described above was placed in a 1-palmitoyl 2-oleoyl phosphatidylcholine (POPC) lipid-bilayer block, of cross-section 75 Å × 75 Å in the xy-plane, with the cytoplasmic side of the protein facing the –z direction. The distance between the layers and surface density of lipid molecules were chosen to match those from experiments and from fully hydrated lipid-bilayer MD simulations. The phosphatidylcholine head group was solvated and some disorder is built into the lipid bilayer patch to obtain a starting structure closer to the real lipid environment. This system was then sandwiched between two blocks of size 75 Å × 75 Å × 15 Å containing pre-equilibrated water molecules. Any lipid and water molecules within 1 Å

and 5 Å respectively of the protein are eliminated. The system is then neutralized by adding the required numbers of sodium or chloride ions, that are placed randomly in the system but avoiding any clashes. This whole setup procedure was implemented using the VMD (Visual Molecular Dynamics) package[47]. The full protein-lipid-water system containing ~37000 atoms was used in the molecular dynamics (MD) simulations discussed next.

F2. Molecular dynamics simulations

Parameters. The simulations were carried out using NAMD[48], a parallelized MD code designed for simulating large biomolecular systems. The CHARMM force field[49] was used for the protein and lipids and the TIP3P potential function[50] was used for the water molecules. Periodic boundary conditions using the Particle-Mesh Ewald summation method[51] with a 12Å cutoff was used for calculating long range interactions. The calculations were performed under isothermal-isobaric conditions (NPT) at 310 K and 1 atm. The temperature is controlled using Langevin dynamics (with a coupling coefficient of 5 ps-1) and the pressure is maintained using a Langevin-Hoover barostat[52]. A time step of 1 fs is used throughout this study.

Simulation protocol: The MD simulations are carried out in 4 steps:

a) The lipid and water atoms were minimized for 10000 steps keeping the protein and ligand atoms fixed. This allows for the lipids and waters to remove any bad contacts with each other and the protein or the ligand.

b) In the second step, the lipid and water atoms were equilibrated under NPT conditions (310K and 1atm) for 500 ps, while keeping the protein and ligand fixed. This lets the lipid and waters equilibrate in the presence of the protein and fill any gaps around the protein created due to system setup.

c) Next, the full system (protein-lipid-water) is minimized for 5000 steps, allowing the protein and ligand to adjust to the equilibrated lipid and waters.

d) In the last step, the full system is equilibrated for 100 ns under NPT conditions, of which the snapshots are saved every 1 ps.

IV. Results and Discussion

A. Bonding Analysis of WT DP-Li/T Receptors at the SuperbiHelix level

Multiple hierarchies of structures demand a reliable and consistent method to evaluate energies. We ranked the structures according total neutral energies as discussed above. As mentioned above, the most stable receptor has a 1-2-7 inter-helical hydrogen bonding network ($N34^{1.50}$ – $K76^{2.54}$ – $S316^{7.46}$), where lysine is unique to the DP receptor in the prostaglandin family. This structure also has a 1-2 $N34^{1.50}$ – $D72^{2.50}$ interaction (Figure 5), where asparagine and aspartic acid are highly conserved in Class A and prostaglandin receptors. After docking and optimizing the binding site, we find an alternate 1-2-7 interhelical bonding network $N34^{1.50}$ – $D72^{2.50}$ – $S316^{7.46}$, which is then stabilized through the dynamics. Most of the receptors (17/20) have a strong interhelical interaction between $K76^{2.54}$ – $S316^{7.46}$. The $S316^{7.46}$ residue possibly anchors the lysine's position such that the lysine can comfortably make interactions with the ligand during activation.

We find that all twenty of the most energetically favorable WT receptors have strong salt bridges between $K291^{6.61}$ – $D292^{6.62}$, $R284^{6.54}$ - $D305^{7.35}$, $E304^{7.34}$ - $R307^{7.37}$, and $D72^{2.50}$ – $K76^{2.54}$. The TM6, TM7 salt bridges probably provide stability to the extra-cellular end of TM6 and TM7 as the receptor undergoes conformational changes. Many other interhelical interactions are found in the top 20 energetically favorable receptors. For instance, $Q122^{3.42}$ and $D319^{7.49}$ interact in nine of the twenty lowest energy receptors. This is of relevance since $Q122^{3.42}$ is unique to the DP receptor, and it has been implicated in stabilizing the in-active conformation in our previous DP MembStruk model[15].

As briefly mentioned, the 3D structure of the bovine rhodopsin GPCR, contained an ionic lock between Arg on TM3 of the "DRY" motif and the adjacent Glu on TM6 that is widely believed to be associated with the inactive GPCR. Thus activation of rhodopsin led to the disruption of this ionic lock. In contrast to most Class A rhodopsin GPCRs, DP receptor does not have the DRY motif. Instead, the 97% conserved Arg^{8} at the second position of the DRY motif is substituted with Cys for the DP receptor. Indeed, we do not find a 3-6 ionic lock in the top WT DP receptor. However, we find two other relevant interactions in five of the top 20 WT receptors, namely $E129^{3.49}$ – $R230^{5.62}$ and $E129^{3.49}$– $H263^{6.33}$. Although these interhelical contacts are viable, they do not indicate stability during our dynamics trajectory.

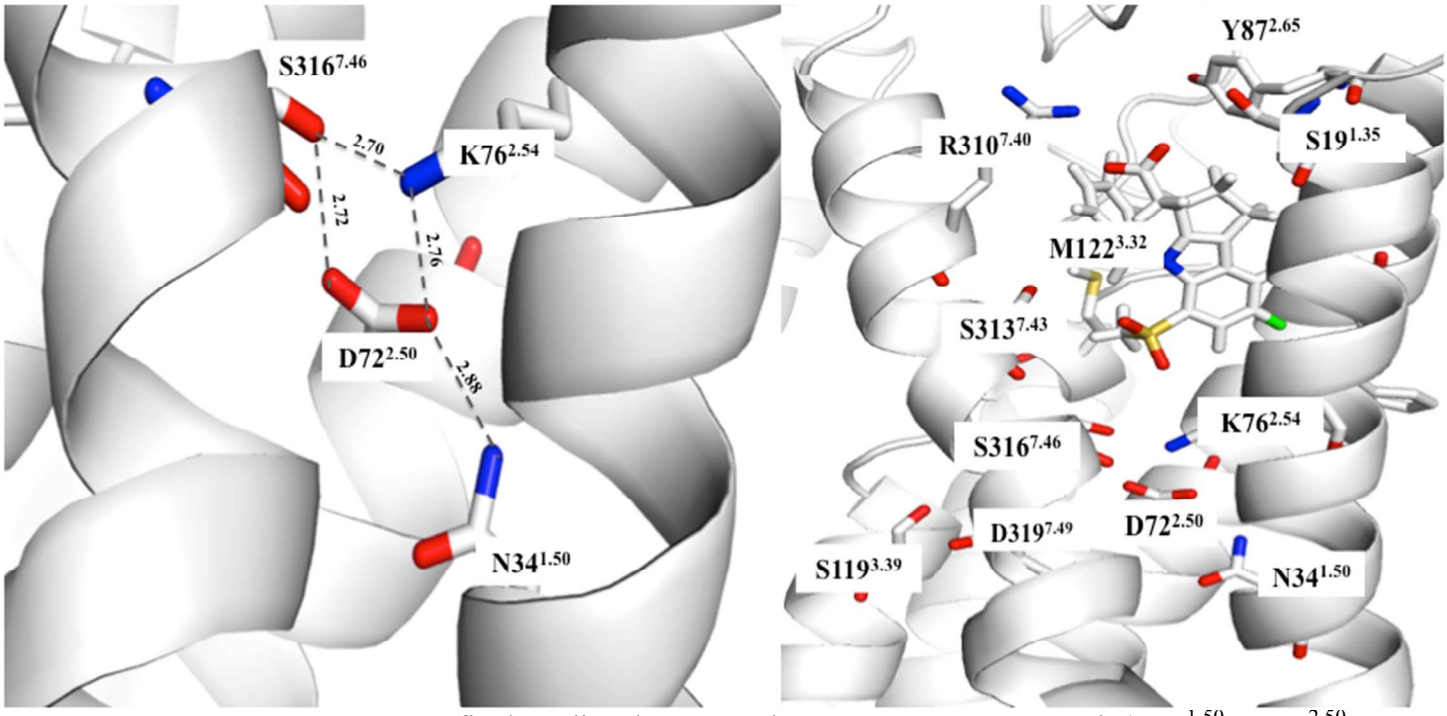


Figure 5. Top Left: Our final predicted structure has a strong 1-2-7 network ($N34^{1.50}$ – $D72^{2.50}$ – $K76^{2.54}$ – $S316^{7.46}$), where N1.50, D2.50 are conserved in Class A GPCRs. The K2.54 is unique to the DP receptor, while the S7.46 is found only in hDP and hE2R2. **Top Right**: We find in our final structure the ligand pocket interacts most prominently with $R310^{7.40}$, $K76^{2.54}$, $S19^{1.35}$. Some of the neighboring residues in the ligand-binding pocket have also been marked, namely

M112$^{3.32}$, S313$^{7.46}$. We have also marked S119$^{3.39}$ – D319$^{7.49}$ interaction, a hydrogen bond stabilized during MD that is implicated in activation.

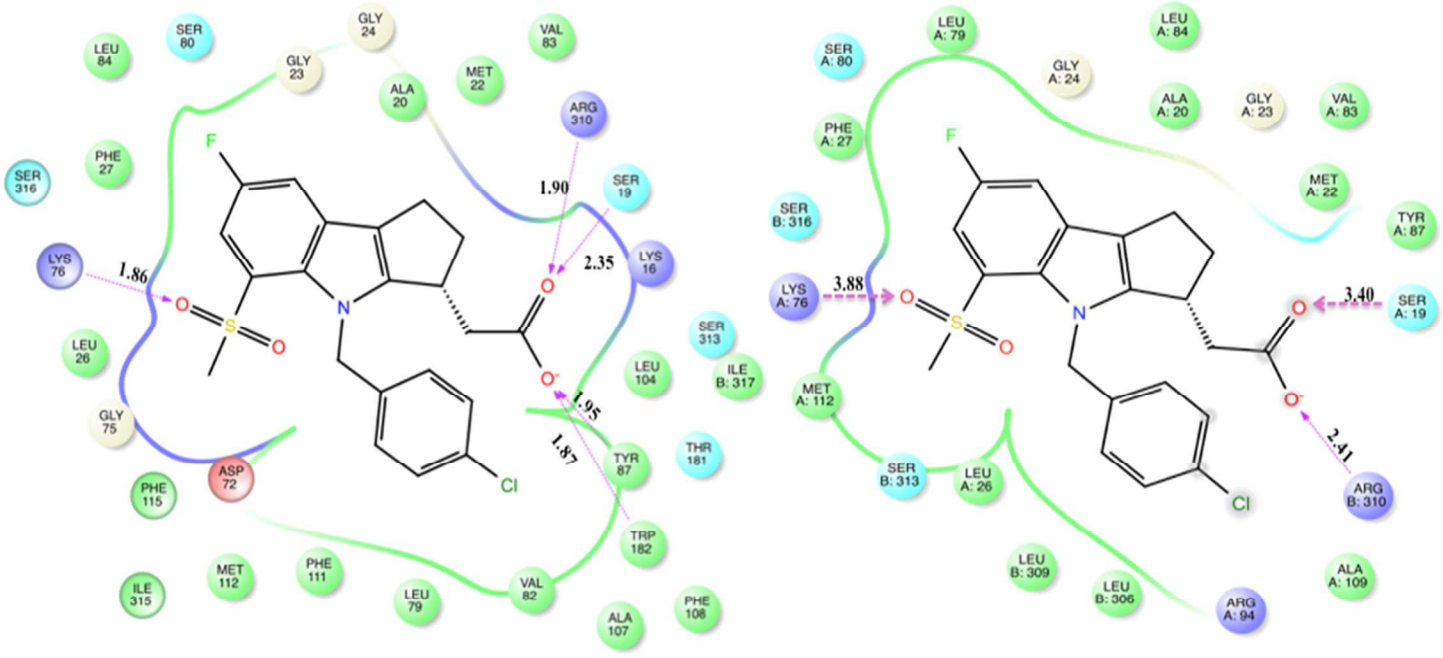


Figure 6. Left: The final predicted binding mode from our previous Membstruk[15] predictions. **Right**: Predicted binding mode selected from our current MD trajectory. The pharmacophore arrangement does not capture the water-mediated interactions between the receptor and the ligand. This snapshot shows R7.40 interacting strongly with the ligand, which is consistent with the experimental observation that R7.40 is strongly implicated in ligand-binding.

As mentioned above, we also find a particularly strong salt-bridge between D72$^{2.50}$ - K76$^{2.54}$ in the middle of TM2. Previously, it has been shown in the A_{2A}AR receptor[29] that Na^+ allosterically binds to D72$^{2.50}$ and S119$^{3.39}$, conserved in most Class A GPCRs. We find an alternate compelling explanation for the behavior of these residues in the hDP receptor. In hDP receptor, we find a stable salt-bridge between D72$^{2.50}$ - K76$^{2.54}$ and another stable interhelical bond between D319$^{7.49}$– S119$^{3.39}$/Q122$^{3.42}$. Since the K76$^{2.54}$ residue is unique to DP receptor and S119$^{3.39}$ is semi-conserved in the prostaglandin family (found only in hDP, hIP, and hE2R3), we

hypothesize that hDP receptor is the only receptor in prostaglandin family that has both of these interactions.

Compared to the previous DP MembStruk model[15], we find that our most favorable energetic complex shares the same 1-2-7 interaction as well specific 2-7 interactions involving N34$^{1.50}$, D72$^{2.50}$, K76$^{2.54}$, and S316$^{7.46}$. However, we find great improvement in our structure, particularly in TM4 and TM5, due to numerous new stabilizing interactions, including G220$^{5.52}$ - N224$^{5.56}$, P167$^{4.60}$ - Y203$^{5.36}$, T214$^{5.46}$ - N218$^{5.50}$. These interactions give our structure greater helical stability, especially in the inactive conformation.

B. Selection of Ligand

We docked the DP prostaglandin receptor to the cyclopentanoindole (CPI) Merck ligand ((−)-[(3*R*)-4-(4-chlorobenzyl)-7-fluoro-5-(methylsulfonyl)-1,2,3,4-tetrahydrocyclopenta[b]indol-3-yl]acetic acid) (Figure 9). We chose this ligand choice because Merck later published detailed SAR studies, which allows us to assess the accuracy of the predicted structure. Additionally, recent studies have also characterized the cyclopentanoindole (CPI) Merck ligand as the first known inverse agonist[16]. The CPI ligand displays pharmacochaperone activity towards the DP receptor[16], making it an interesting ligand case to study.

C. Final Predicted Binding Mode

After docking the ligand to our top 20 receptors (based on the procedure above), we ranked the complexes according to complex total energy (Table 2). We compared the complexes based on complex total energy, since it allowed for a consistent comparison of complexes of different proteins from our WT hierarchy docked to the same ligand. We then optimized the top complex (as described above) to find our lowest energy SCREAMed complex.

We discuss the binding of the 4 regions of the CPI Merck ligand (Fig. 9) with the receptor residues: (1) carboxylic acid (2) sulfonyl group (3) benzene ring (4) indole ring. We have also indicated the total energy contribution of the residue interaction with the CPI antagonist.

(1) **Carboxylic acid**: We find two strong interactions of the carboxylic acid with R310$^{7.40}$ (-25.986 kcal/mol) and Y87$^{2.65}$ (-4.460 kcal/mol) on TM7 and TM2 respectively.

(2) **Sulfonyl group:** We find the sulfonyl group forms a hydrogen bond with K76$^{2.54}$ (-19.618 kcal/mol) on TM2. The sulfonyl group also has weaker interactions with M112$^{3.32}$ (-0.148 kcal/mol) on TM3.

Top 20 Antagonist-Receptor Complexes from Docking CPI Antagonist to Top Ranked SuperBiHelix Structures Based on Neutral Total Energies

Rank	WT Receptor	Complex Total Energy	Snap Binding Energy
1	**1**	**-350.355**	**-56.932**
2	1	-333.706	-53.726
3	1	-328.896	-45.712
4	1	-327.304	-46.562
5	1	-325.424	-47.122
6	9	-321.794	-49.971
7	1	-316.730	-50.203
8	6	-311.073	-52.798
9	1	-309.932	-48.027
10	1	-308.478	-48.528
11	1	-305.772	-46.135
12	14	-303.316	-53.360
13	6	-303.278	-50.381
14	1	-302.680	-48.189
15	14	-302.399	-53.725
16	9	-301.281	-49.251
17	14	-297.638	-52.620
18	14	-297.044	-53.203
19	14	-295.077	-56.505
20	14	-291.569	-53.700

Table 2. 20 most energetically favorable antagonist-receptor complexes based on complex total energy.

(3) **Benzene ring**: We find the benzene ring is located between TM2 and TM3, interacting with residues such as V82$^{2.60}$ (-1.909 kcal/mol).

(4) **Indole ring:** We find the indole ring interacts with residues on TM1 and TM2, including S80$^{2.58}$ (-1.961 kcal/mol), G23$^{1.39}$ (-0.447 kcal/mol), L84 (-1.764 kcal/mol).

D. Comparison of Predicted Binding Mode to Mutagenesis Data:

Previous SAR predictions led to a 1000 fold improvement in binding over the original lead compound (K_i: 800nM to 0.8 nM). Our predicted binding mode of the parent CPI (Figure 6) is in good agreement with experimental – site-directed mutagenesis data. There is little site-directed mutagenesis data on the DP receptor; however, we can infer the function and

importance of critical amino acid residues based on the analogous residues on other prostaglandin receptors.

Binding Site: The location of our binding pocket agrees with experimental data by Kobayashi et. al[17] and the structure predictions of Li et al[15], who concluded that the ligand binding pocket for prostaglandin receptors was formed by the TM1, TM2, and TM7 regions. They found that both TM-1 and TM-2 are involved in recognition of the ring component of the prostaglandin lipid compound while TM-7 is involved in recognition of the side chains. Li et. al[15] also docked to the PGD_2 agonist, finding strong interactions between the agonist and TM2 and TM7. Most notably, they found that the agonist forms strong hydrogen bonds between 9-OH of the cyclopentane ring and $S313^{7.43}$, 15-OH of the ω chain and $S316^{7.46}$, and a strong salt-bridge interaction between 1-COOH of the α chain and $R310^{7.40}$. Additionally, Li et. al[15] found that the agonist binding disrupts the $K76^{2.54}$-$S316^{7.46}$ hydrogen bond found in the apo structure, leading to the clockwise rotation of TM3 and counterclockwise rotation of TM7, which is probably characteristic of DP activation. While we did not dock our final selected receptor to PGD_2 agonist, we find a similar binding region for the antagonist binding pocket, where TM2 and TM7 residues (e.g. - $R310^{7.40}$, $K76^{2.54}$, $Y87^{2.65}$) bind strongly to the antagonist on the extra-cellular end of the receptor (Figure 6). This is consistent with studies done by Stitham et. al on hIP receptors[26,27], which showed the binding pocket for the receptor is in the upper-half of the TM region.

TM-2: We find that $K76^{2.54}$ binds strongly (~-20 kcal/mol) to the sulfonyl group and $Y87^{2.65}$ binds (~-4.5 kcal/mol) to the carboxylate group on the antagonist. Kobayashi concluded that $K76^{2.54}$ on TM-2[17] is required for high affinity DP binding to the PGD2 endogenous ligand. We find a strong interaction between $K76^{2.54}$ and the CPI antagonist, which implies that this lysine is also required for high affinity DP binding to antagonist. Stitham et. al[18] found that tyrosine at

the extra-cellular end of TM2 made up the immediate binding pocket cluster in the hIP receptor. The tyrosine on TM2 is conserved throughout the majority of the prostaglandin receptors, except for thromboxane receptors. The analogous tyrosine on the IP receptor $Y87^{2.65}$ was shown to be essential for proper receptor activation, when expression levels were measured experimentally. The interactions of the antagonist in our predicted binding mode with $K76^{2.54}$ and $Y87^{2.65}$ is therefore consistent with the experimental data.

TM-3: We find weaker binding of the antagonist $M112^{3.32}$ (-0.148 kcal/mol) on TM3, where $M112^{3.32}$ is universally conserved in the prostaglandin receptor family, implying its importance in receptor activation and function. Therefore, our predicted binding mode is consistent with experimental data from Stitham and Bell et. al[30,31], who found that mutating M3.32 reduced ligand-binding and even protein-misfolding[30].

TM-7: We find a very strong interaction between the carboxylate on the antagonist and $R310^{7.40}$ (-25.986 kcal/mol) on TM7. This interaction between antagonist and $R310^{7.40}$ seems likely, given the mutation data for the prostagladin receptors. Although no experimental data has been reported regarding the importance of $R310^{7.40}$ for the DP receptor, this arginine is universally conserved and has been experimentally shown to be critical for interacting with the carboxylic acid on the Prostaglandin D_2 ligand on the IP[18], EP2[19], EP3[20], FP[21], and TP[22] receptors. The asparagine (N7.49) on TM7 is typically involved in the 1-2-7 conserved interhelical interactions. This NPxxY structural motif seen in more than 95%[23] of Class A GPCRs is *not found* in the prostaglandin DP receptor. The experimental analysis by Audoly and Breyer[24] found that this aspartic acid ($D319^{7.49}$) on TM7 does *not* participate directly in the receptor-ligand binding but can provide the receptor flexibility to undergo conformational changes. This is consistent with our predicted binding mode, where we find that $D319^{7.49}$ does not directly participate in ligand

binding, but rather forms a strong hydrogen bond interaction forms during MD between S119$^{3.39}$ - D319$^{7.49}$ (Figure 6, Figure 9). As suggested by previous agonist and antagonist studies on DP[15], this could imply that movement of D319$^{7.49}$ or displacement of TM7 with respect to TM3 is characteristic of receptor activation, where the inactive conformation has a strong hydrogen bond S119$^{3.39}$ - D319$^{7.49}$.

E. Stability of Interactions During Molecular Dynamics:

E.1 interhelical hydrogen bonds during MD. During the MD, we observed the following (Figure 7):

- **1-2-7 Interaction:** Initially, the hDP receptor has the N34$^{1.50}$- D72$^{2.50}$ – S316$^{7.46}$ 1-2-7 hydrogen bonding network, where the N34$^{1.50}$- D72$^{2.50}$ interaction bond distance varies between 2.0 – 5.4 Å. This bond is further stabilized after 10 ns in MD, as the distance between the closest oxygen on aspartic acid and nitrogen on asparagine's side chain remains between 1.6 – 2.7 Å from 10-100 ns. After 10 ns, there is also a slight increase in the bond distance between the closest oxygen on D72$^{2.50}$ and S316$^{7.46}$ (Figure 7), where the bond distance varies from 2.8 – 4.8 Å in the first 10 ns and 3.1 to 5.2 Å from 10 – 100 ns. There is also little variation in the closest bond distance between D72$^{2.50}$ and K76$^{2.54}$, which remains between 2.4 – 3.1 Å through all 100 ns. As the average bond distance between the closest heavy atoms D72$^{2.50}$ and S316$^{7.46}$ increases from 3.9 (0-10 ns) to 4.3 Å (10-100 ns), several other favorable interactions can account for this increase. Since D72$^{2.50}$ is anchored by favorable interactions with both N34$^{1.50}$ and K76$^{2.54}$, it is unable to move closer to S316$^{7.46}$, which forms new interactions with the sulfonyl group on the ligand (Fig. 8, Section E.2**).** Additionally, there is little change in the bond distance between K76$^{2.54}$ – S316$^{7.46}$, which remains stable between 2.5 to 3.5 Å through all 100 ns of MD. Therefore, in addition to the

key interactions with $K76^{2.54}$, namely $K76^{2.54}$ – $S316^{7.46}$ and $D72^{2.50}$ – $K76^{2.54}$, we find that the 1-2-7 hydrogen bonding network, $N34^{1.50}$- $D72^{2.50}$ – $S316^{7.46}$, predicted in our SuperBiHelix structures, remains stable throughout 100 ns of MD. The main difference in these interactions before and after 10 ns is the slight increase in bond distance between $D72^{2.50}$ and $S316^{7.46}$, which can be explained by the new interaction of $S316^{7.46}$ with the CPI ligand.

- **3-7:** Although our initial docked pose lacked the $S119^{3.39}$ - $D319^{7.49}$ interaction, we find this interaction stabilizes immediately during MD. In the first 10 ns of MD, we find the bond fluctuations decrease, as the distance drops from 4.5 Å to remain within 2.4-4.2 Å. The bond distance was measured by calculating the minimum distance in each frame between the closest oxygen on aspartic acid's side chain and oxygen on serine's side chain (two closest heavy atoms). In our extended 100 ns MD simulation, this bond stabilizes as the bond distance remains between 2.4 Å to 3.1 Å (Figure 7). Especially since prostaglandin receptors have a unique DPWxF motif and D(7.49) is fully conserved in the prostaglandin family, we expect the aspartic acid to be involved in activation, consistent with mutagenesis data above. The key change is that the $S119^{3.39}$ - $D319^{7.49}$ stabilizes further after 10 ns, indicating that this interaction is lasting in our predicted structures.
- **6-7:** We find that the salt-bridges remain mostly stable between TM6 – TM7 during 100 ns of MD. In addition to the salt bridge between $R307^{7.37}$ - $E304^{7.34}$ (residues unique to hDP receptor), we find a salt-bridge network between $K291^{6.61}$ – $D292^{6.62}$ - $R284^{6.54}$ - $D305^{7.35}$. Specifically, we find when the receptor loses $K291^{6.61}$ – $D292^{6.62}$ during MD (e.g. from 25-55 ns), the receptor forms a new strong salt-bridge interactions with bond distance 2.5-3 Å between $D292^{6.62}$ - $R284^{6.54}$ and $D292^{6.62}$ - $E304^{7.34}$. Our final selected binding pose has the

following salt-bridge network: D292$^{6.62}$ - R284$^{6.54}$ - D305$^{7.35}$, in addition to the R307$^{7.37}$ - E304$^{7.34}$ salt-bridge. We hypothesize that these specific salt-bridges are crucial for anchoring TM6 in the inactive conformation, especially since TM6 movement is characteristic of activation, as seen in rhodopsin and β_2-adrenergic receptors[34].

Therefore, compared to the SuperBihelix structures, we note that the protein-protein hDP interhelical bonds stabilize throughout MD. The key 1-2-7 interhelical bonds either grow stronger (e.g. N34$^{1.50}$- D72$^{2.50}$) or remain stable (e.g. D72$^{2.50}$ - S316$^{7.46}$, K76$^{2.54}$ - S316$^{7.46}$) throughout our extended 100 ns MD trajectory. This trend also holds when looking at the key interactions between TM3-7 and TM6-7 that are highlighted above. When comparing our MD trajectory from 0 – 10 ns to the extended trajectory from 10 – 100 ns, we note that the interactions continue to stabilize and that the bonds that seemed transient from 0 – 10 ns (e.g. S119$^{3.39}$ – D319$^{7.49}$), remain stable in the extended trajectory.

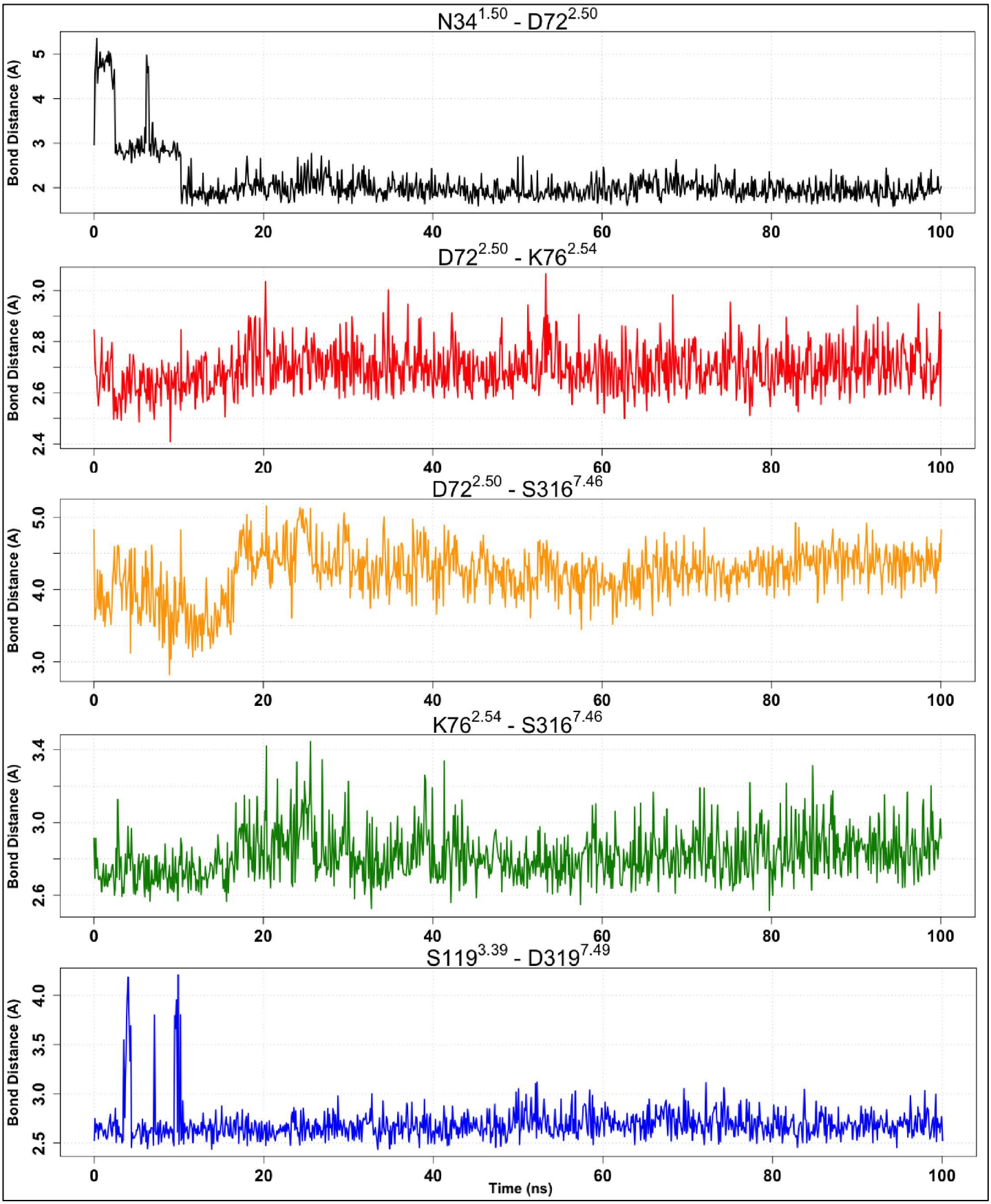
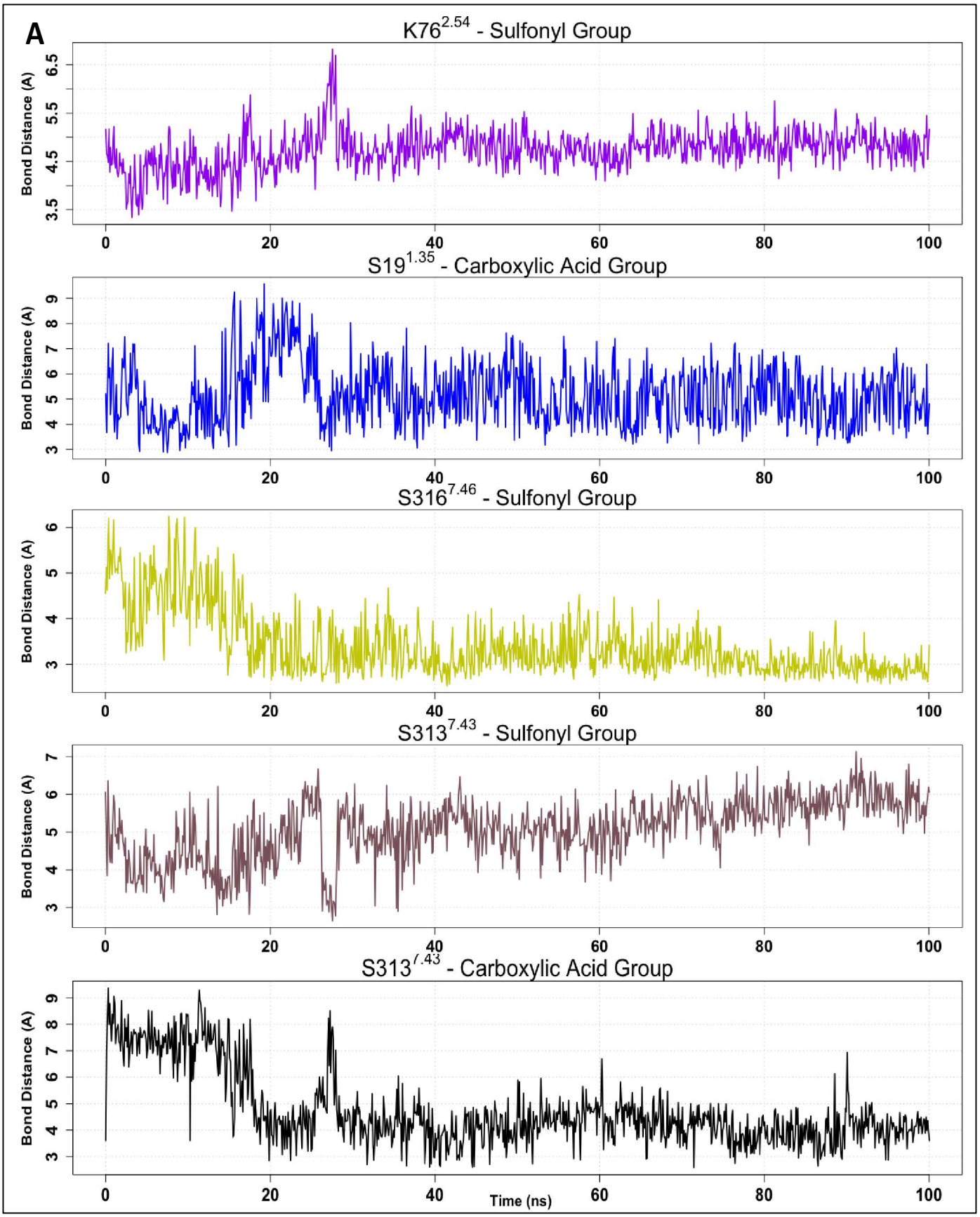
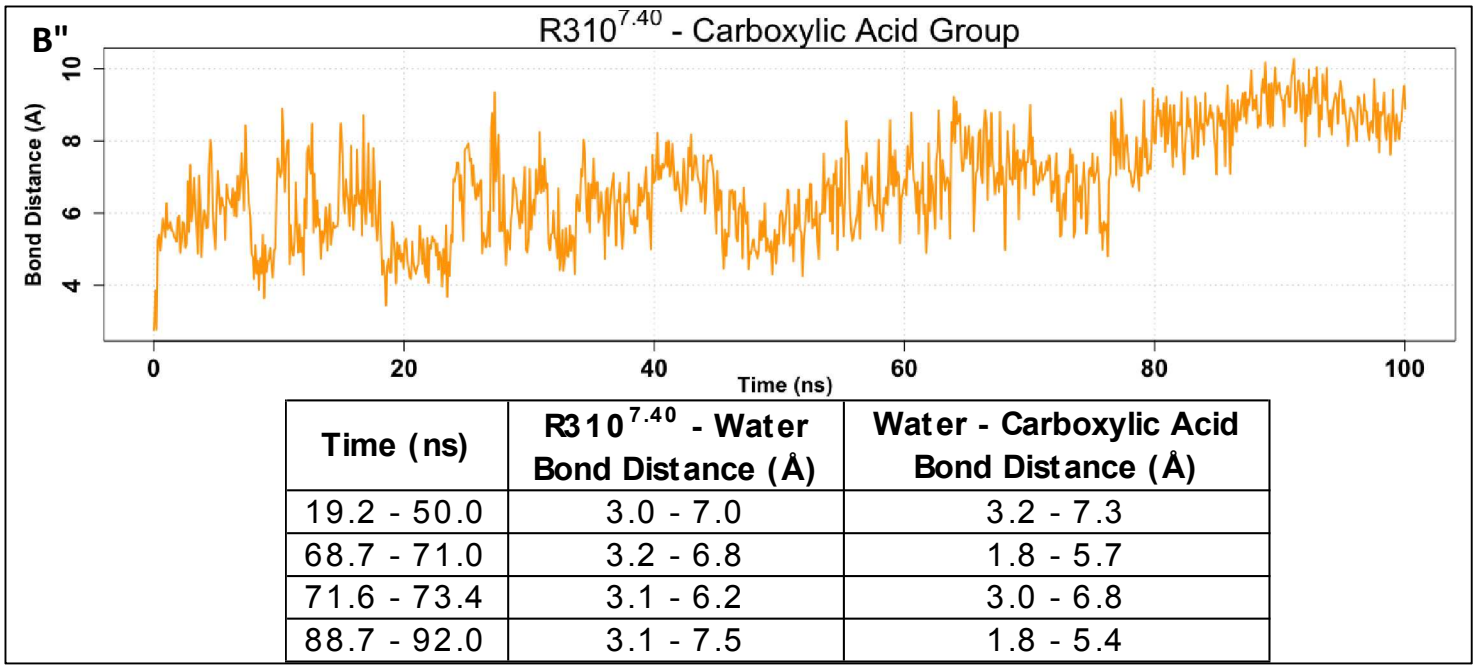

N34^1.50 - D72^2.50
D72^2.50 - K76^2.54
D72^2.50 - S316^7.46
K76^2.54 - S316^7.46
S119^3.39 - D319^7.49
Bond Distance (A)
Time (ns)

Figure 7. Trajectory analysis of hydrogen bonding among TM helices during 100 ns MD. The plotted distance is measured as the distance between two closest heavy atoms. We find that the 1-2-7 interactions are mostly conserved during our MD trajectory. N34(1) – D72(2) distance, measured from the closest oxygen on Asp to nitrogen on Asn side chain, drops from 5.4 Å in the first ns to an average of 3.3 Å in the first 10 ns. From 10-100 ns, this bond distance stabilizes further to an average of 2.0 Å. D72(2) – K76(2) distance, measured as the distance between closest oxygen on aspartic acid to nitrogen on lysine's side chain, ranges from 2.4 – 3.1 Å throughout 100 ns. There is little change in bond distance for D72 – K76 throughout MD, as the average bond distance is 2.64 Å from 0 – 10 ns and 2.70 Å from 10 – 100 ns. D72(2) – S316(7) bond distance, calculated as distance between closest oxygen at each frame from oxygen on serine's side chain, ranges from 2.8 to 5.1 Å through 100 ns. The average bond distance slightly increases from 3.9 Å in 0-10 ns to 4.3 Å in 10-100 ns (discussed in E.1). K76(2) – S316(7) bond remains stable as bond distance varies from 2.5 to 3.4 Å, where average bond distance is 2.73 Å from 0 -10 ns and 2.83 Å from 10-100 ns. During MD, we find D319(7) forms a strong interhelical bond with S119(3), where bond distance is measured from the closest oxygen on Asp and oxygen on serine's side chain. This bond distance, which varies from 2.4 – 4.2 Å in 0-10 ns, stabilizes as the bond distance ranges from 2.4 – 3.1 Å from 10-100 ns.


A
K76 2.54 - Sulfonyl Group
S19 1.35 - Carboxylic Acid Group
S316 7.46 - Sulfonyl Group
S313 7.43 - Sulfonyl Group
S313 7.43 - Carboxylic Acid Group
Bond Distance (A)
Time (ns)



Time (ns)	R310 $^{7.40}$ - Water Bond Distance (Å)	Water - Carboxylic Acid Bond Distance (Å)
19.2 - 50.0	3.0 - 7.0	3.2 - 7.3
68.7 - 71.0	3.2 - 6.8	1.8 - 5.7
71.6 - 73.4	3.1 - 6.2	3.0 - 6.8
88.7 - 92.0	3.1 - 7.5	1.8 - 5.4

Figure 8. (A) Key receptor interactions with the CPI ligand stabilize during 100 ns of dynamics. The K76$^{2.54}$ – sulfonyl group bond distance, measured from the closest nitrogen on lysine to closest oxygen on the sulfonyl group, ranges from 3.3 to 6.8 Å. The K76$^{2.54}$ – sulfonyl group average bond distance increases slightly from 4.3 Å from 0 – 10 ns to 4.8 Å in 10 – 100 ns. The distance between side chain oxygen on S19$^{1.35}$ and the closest oxygen on the carboxylic group on the ligand drops at the start of dynamics from 7.2 to 3.8 Å from 0.3 – 5.0 ns, before the bond stabilizes at an average distance of 5.0 Å from 30 – 100 ns. The bond distance between S316$^{7.46}$ and the sulfonyl group, measured from the oxygen on serine's side chain to the closest oxygen on the sulfonyl group, stabilizes after 20 ns, where average bond distance drops from 4.4 Å between 0 – 20 ns to 3.1 Å from 20 – 100 ns. From 0 – 10 ns, S313$^{7.43}$ forms a hydrogen bond with the sulfonyl group with average bond distance of 4.4 Å, when distance is measured between the side-chain oxygen on serine and the closest oxygen on the sulfonyl group. From 10 – 100 ns, S313$^{7.43}$ forms a more favorable contact with the carboxyl group, with average bond distance 4.4 Å from 10 – 100 ns, when measured between closest oxygen on carboxyl group and oxygen on

serine's side chain. As the average bond distance between S313$^{7.43}$ and the carboxyl group drops from 7.4 Å in 0 – 10 ns to 4.4 Å in 10 – 100 ns, the average bond distance increases between S313$^{7.43}$ and the sulfonyl group on the ligand from 4.4 Å in 0 – 10 ns to 5.2 Å in 10 – 100 ns.

(B) Water-Mediated Interaction between R310$^{7.40}$ on receptor and carboxylic group ligand. Bond distance is calculated by finding the minimum distance in each frame between the three nitrogen atoms on arginine's side chain to the closest oxygen on the carboxylic group. The bond distance between R310$^{7.40}$ and the carboxylic group on the ligand varies from 2.8 – 8.9 Å from 0 – 10 ns and 3.4 – 10.3 Å from 10 – 100 ns. Although the distance between R310$^{7.40}$ and the carboxylic group ligand increases throughout MD from 5.7 Å in 0 – 10 ns to 6.9 Å in 10 – 100 ns, this interaction is still present and mediated by several water molecules that diffuse into the binding pocket. We have noted the time frames for four water molecules that mediate this interaction throughout MD. Bond distance was measured from closest nitrogen on arginine's side chain to closest oxygen on the water molecule. For instance, between 19.2 – 50.0 ns, R310$^{7.40}$ forms a stable interaction with a water molecule with bond distance between 3.0 – 7.0 Å, before the water molecule diffuses out of the pocket.

E.2 receptor-ligand interactions. During MD, we observed the following (Figure 5, Figure 6, Figure 7):

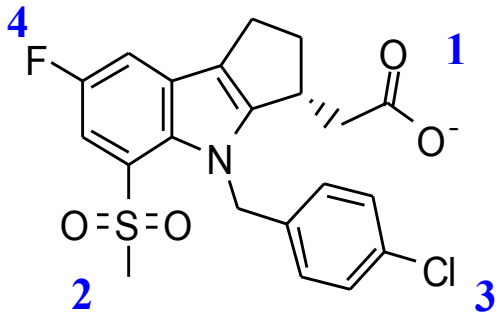


Figure 9. Cyclopentanoindole (CPI) Merck ligand Regions: 1: Carboxylic Acid, 2: Sulfonyate group, 3: Benzene Ring, 4: Indole Ring

(1) Carboxylic Acid: We find an interaction between carboxylic acid and $R310^{7.40}$ that ranges from 2.7 to 10.3 Å (Figure 8B). Initially, in the first 10 ns of MD, we can explain the high variability in bond distance between R7.40 and the ligand ranging from 2.8 to 8.9 Å, by noting the interaction between $R310^{7.40}$ and $L306^{7.36}$, which restricts the ability for arginine to form the salt-bridge with the ligand. For example, as the ligand-receptor complex equilibrates, we find that $R310^{7.40}$ interaction with $L306^{7.36}$ stabilizes to within 3 – 7.5 Å, which moves the arginine farther from the carboxylic group. At 9.5 ns, the bond distance between $R310^{7.40}$ and the carboxylic acid group drops to 4.2 Å. As MD is extended from 10 – 100 ns, $R310^{7.40}$ is still able to interact with the carboxylic acid, as four water molecules during 19.2 – 50, 68.7 – 71.0, 71.6 – 73.4, 88.7 – 92.0 ns mediate this interaction (Figure 8B). Therefore, although the measured distance between $R310^{7.40}$ and the closest oxygen on the carboxyl group grows larger from 5.8 Å in 0 – 10 ns to 6.9 Å in 10 – 100 ns, this difference can be explained as the arginine residue is still able to interact with the carboxyl group on the ligand through a water-mediated interaction. We also find at the very start of MD a hydrogen bond forms between $S19^{1.35}$ and the carboxylic acid (Fig. 8A) on the ligand that stabilizes from 30 – 100 ns with an average bond distance at 5.0 Å. The spike in bond distance at around 20 ns can be explained in that $S19^{1.35}$ temporarily forms a bond with $Y87^{2.65}$, before the bond between S19 and the ligand stabilizes through the rest of the 100 ns. This interaction is of significance since $S19^{1.35}$ is almost fully conserved in the prostaglandin receptor family. The other key change in protein-ligand interactions is how $S313^{7.43}$ interacts with the ligand. As noted in Fig. 8, initially, $S313^{7.43}$, which is semi-conserved in the prostaglandin family, forms a hydrogen bond with the sulfonyl group with 4.4 Å from 0 –

10 ns. As the complex further relaxes in MD, at 19.4 ns, S313$^{7.43}$ finds a new bond with the carboxylic group (Fig. 8A) as the bond distance drops to 3.5 Å, with this interaction stabilizing at around 5.2 Å throughout the rest of MD.

(2) Sulfonyl Group: We find the interaction between K76$^{2.54}$ and the sulfonyl group on the ligand (Fig. 8A), found in our DarwinDock structures, stabilizes with bond distance ranging from 3.3 to 6.8 Å. As mentioned in Fig. 8A, we note that the average bond distance between K76$^{2.54}$ and the sulfonyl group slightly increases from 4.3 Å in 0-10 ns to 4.8 Å in 10 – 100 ns. The main reason for this slight increase is that the lysine is optimizing its interactions with D72$^{2.50}$ and S316$^{7.46}$, which forms a new bond with the sulfonyl group on the ligand that stabilizes between 18-20 ns. As the S316$^{7.46}$ bond distance drops with the sulfonyl group on the ligand, the lysine is constrained by its interaction with S316$^{7.46}$, which is reflected in the slight increase in the K76$^{2.54}$ - the sulfonyl group bond distance. This is the one of the largest differences in the MD between 0 – 10 ns and 10 – 100 ns, as the distance between the side chain oxygen on S316$^{7.46}$ and the closest oxygen in the sulfonyl group drops from 5.0 – 3.5 Å from 10.4 - 19.8 ns. This bond stabilizes through the rest of MD, with a mean distance of 4.7 Å. The strength and reduced bond distance of this interaction also explains why the D72$^{2.50}$ – S316$^{7.46}$ bond grows weaker during MD, since the serine is anchored by favorable interactions with K2.54 and the sulfonyl group. Additionally, we also find the sulfonyl group forms stable contacts with F27$^{1.43}$, M112$^{3.32}$, ranging from 3.5-6.5 Å, when measured from the closest oxygen on the sulfonyl group to the sulfur on methionine's side chain, which is seen throughout the 100 ns of MD. These interactions are of relevance to other prostaglandin receptors, since F1.43 is semi-conserved and M3.32 is almost fully conserved in the prostaglandin receptor family.

(3) Benzene Ring: The benzene ring is located between TM3 and TM7. The benzene ring is surrounded by F108$^{3.28}$, A109$^{3.29}$, and L309$^{7.39}$, where F3.28 and A3.29 are semi-conserved in the prostaglandin family. These interactions are also consisted with previously published binding modes between CPI antagonist and DP[15]. Also, we find that Y2.65 replaces its initial interaction with the carboxylic acid before dynamics with an interaction with chlorine, during the first 25 seconds of MD. The tyrosine interaction with the ligand alternates during the MD between interactions with the closest oxygen on the carboxylic group and alternate favorable interactions with S19$^{1.35}$ and D101 on the EC2 loop region.

(4) Indole Ring: The indole ring stabilizes in the 1-2-7 binding pocket and interacts with F27$^{1.43}$, L79$^{2.59}$, and S80$^{2.60}$. Within these residues, both L2.59 is unique to hDP, while F1.43 and S2.60 are semi-conserved in the prostaglandin family. Thus, we expect that F1.43 and S2.60 will be part of the binding pocket for other receptors in the prostaglandin family.

In summary, we note that the receptor-ligand binding of our docked structures improves through MD. Below we have summarized how MD improves the binding pose of our DarwinDock structures:

- **0 – 10 ns:** In our selected docked pose, we note interactions between R310$^{7.40}$, Y87$^{2.65}$ with the carboxyl group and K76$^{2.54}$, M112$^{3.32}$ with the sulfonyl group on the ligand. In MD, R310$^{7.40}$ bond distance with the carboxyl group varies from 2.8 – 8.9 Å, which can be explained by arginine's interaction with L306$^{7.36}$ from 7 – 10 ns. Y87$^{2.65}$ bond distance varies from 3.2 – 6.5 Å, which can be explained by the fact Y87$^{2.65}$ optimizes its hydrogen bond with S19$^{1.35}$ and moves closer to the benzene ring (distance measured from chlorine to side chain oxygen on tyrosine) to below 3.2 Å. In the first 10 ns of MD, we note that K76$^{2.54}$ interactions with the sulfonyl group stabilize with an average bond

distance of 4.3 Å. As the bond distance between M112$^{3.32}$ and the sulfonyl group starts at 5.1 Å, this bond distance, measured between sulfur on methionine's side chain with the closest oxygen on the sulfonyl group, fluctuates from 4.0 to 9.4 Å in the first 10 ns. This fluctuation can be explained by the fact that S316$^{7.46}$ moves closer to the sulfonyl group from 5 – 10 ns with average bond distance of 4.7 Å, which alters M112$^{3.32}$ bond distance with the sulfonyl group to minimize any steric clash. The two other key improvements in the receptor-ligand interaction is the formation of new hydrogen bonds that form and stabilize during the first ten ns of MD, including S313$^{7.43}$ and the sulfonyl group with an average bond distance of 4.4 Å and S19$^{1.35}$ and the carboxyl group with average bond distance 4.4 Å.

- **10 – 100 ns:** In extending the MD from 10 – 100 ns, we find a couple key changes that further stabilize the protein – ligand binding. First, we find S316$^{7.46}$ forms a new hydrogen bond with the sulfonyl group on the ligand at around 20 ns, which maintains an average bond distance of 3.2 Å from 20 – 100 ns and further improves the favorability of the binding pose. Second, we find that the hydrogen bond that formed between S313$^{7.43}$ and the sulfonyl group in the first 10 ns of MD grows weaker as the average bond distance increases from 4.4 in 0 – 10 compared to 5.1 Å in 10 – 100 ns. Instead, we find that S313$^{7.43}$ forms a different hydrogen bond with the carboxyl group that stabilizes from 30 – 100 ns, with average bond distance of 4.1 Å. In comparing the other receptor-ligand interactions, we find that although the physical distance between R310$^{7.40}$ and the carboxyl group on the ligand grows larger, R310$^{7.40}$ still interacts with the carboxyl group through four water molecules that diffuse in and out of the pocket through the extended MD trajectory (Fig. 8B). We note the newly formed hydrogen bond between S19$^{1.35}$ and

the carboxyl group from 0 – 10 ns is preserved from 10 – 100 ns, as it maintains an average bond distance of around 5 Å. The variability in bond distance for S1.35 can be explained by transient hydrogen bond interactions with Y87$^{2.65}$. The Y87$^{2.65}$ interaction with the carboxyl group on the ligand ranges from 3.5 to 8.5 Å and is maintained for parts of the extended trajectory. The major sources of variability for this interaction comes from the formation of transient hydrogen bonds with S1.35 and other interactions it forms with D101 on the EC2 loop region. The extended MD from 10 – 100 ns also preserves the hydrogen bond between K76$^{2.54}$ and the sulfonyl group, which maintains an average bond distance of 4.8 Å. Additionally, M112$^{3.32}$, which optimizes its position in the binding pocket to minimize steric clash with S316$^{7.46}$, maintains a bond distance of 4.8 Å with the sulfonyl group on the ligand.

Therefore, we find that MD indeed improves the ligand-receptor interactions on TM 2, 3, 7 by stabilizing the existing interactions from the DarwinDock level including R310$^{7.40}$, Y87$^{2.65}$, K76$^{2.54}$, M112$^{3.32}$. Additionally, relaxing the receptor-ligand complex allows for the formation of multiple new favorable interactions including S19$^{1.35}$, S313$^{7.43}$, S316$^{7.46}$.

F. Comparison of Predicted Binding energies to Experimental Binding constants

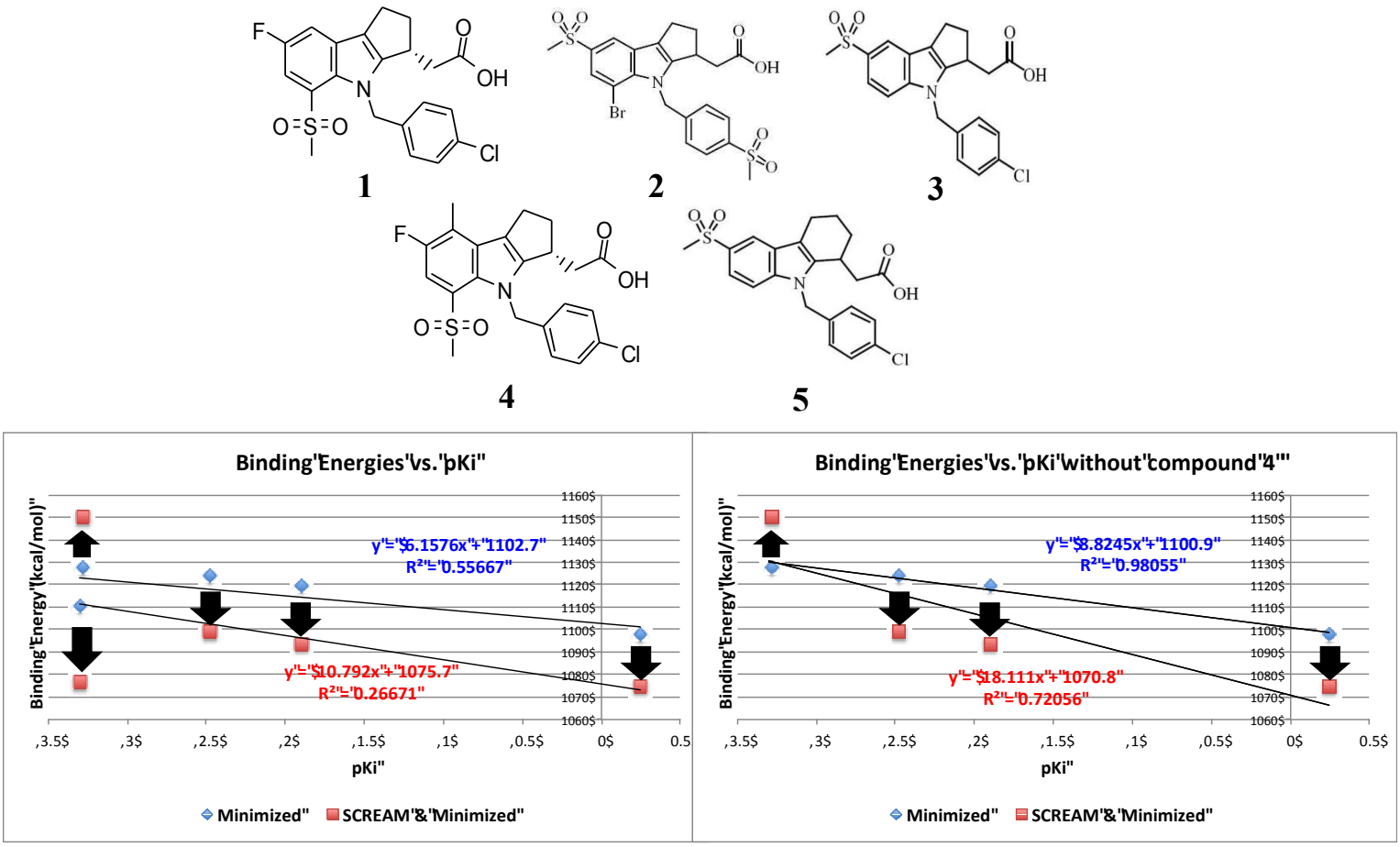


Figure 10. Predicted binding energies of cyclopentanoindole (CPI) and 4 derivatives, compared to the pKi values reported by Merck[32,33].

We find reasonable agreement between predicted binding energies (complex total energy) and experimental pKi values (Figure 10), when we docked the antagonist variants to the same position as 1 and minimized the bind site to one of the snapshots selected from MD.

1: When we SCREAM[14] and minimize the binding site, we find the first compound-receptor binding is most favorable, similar to experimental data (compound 1, Figure 10). Most prominently, we find strong interactions between the compound 1 and K76$^{2.54}$ (TM2) and R310$^{7.40}$ (TM7) on the receptor. Our calculations predict that binding energies for the first and fourth compound are roughly comparable (ΔBinding Energy = 2.2 kcal/mol) after SCREAM and

minimize, which is not consistent with the experimental data. We discussed below this disparity in our description of compound 4 binding.

2: Comparing just the minimized binding site, we find that this binding pose has unfavorable contacts with $F27^{1.43}$, $S80^{2.58}$, $L306^{7.36}$ contributing to a total increase in binding energy by ~30 kcal/mol compared to the minimized complex with compound 1. When we SCREAM and minimize the binding site, we find a combined improvement of ~4.6 kcal/mol in the contacts between compound 2 and $F27^{1.43}$, $S80^{2.58}$, $L306^{7.36}$ on the receptor compared to the minimized compound 2 receptor complex. This is the only antagonist-receptor complex, where our SCREAM and minimize procedure increased the complex total energy compared to the complex total binding energy from just minimizing the binding site. Part of this difference is explained from the change in binding energies between the minimized and SCREAM/minimized antagonist and the binding pocket residues on TM7. Particularly, SCREAMing on the binding pocket, improves interactions with $L306^{7.36}$, $L309^{7.37}$, $S313^{7.43}$, $S316^{7.46}$, $I317^{7.47}$, at the cost of weakening the strong hydrogen bond interaction between $R310^{7.40}$ for a change in binding energy by ~8 kcal/mol.

3: After SCREAMing and minimizing the binding site, we find this binding pose loses its strong interactions between the ligand and $K76^{2.54}$, $S313^{7.43}$, $S316^{7.46}$ on the receptor, thereby contributing to a total predicted decrease in binding energy (compared to compound 1) by 24.4 kcal/mol. We find that the calculated decrease in binding is greater than the experimental decrease by a factor 526. A possible reason for this difference is that we did not re-dock the ligand so that the optimum-binding pose may change.

4: This compound leads to the biggest inconsistency between experimental and predict binding mode. Comparing the binding energies after *just* minimization between compound 1 and

compound 4 receptor complex, we find a decrease in predicted binding of 12.7 kcal/mol compared to compound 1, because of the clash of the methyl group on compound 4 with V83$^{2.61}$ on the receptor (~9.2 kcal/mol). But SCREAMing[14] and minimizing the complex with compound 4 mostly resolves this clash, which is expected since SCREAM optimizes rotamers with the worst clashes. Overall, this yields a complex compound 4 after SCREAMing and minimization is 2.3 kcal/mol less favorable in binding energy compared to the corresponding complex with compound 1.

5: After SCREAMing and minimization this binding pose led to unfavorable contacts with S80$^{2.58}$, while also losing favorable interactions with K76$^{2.54}$, S313$^{7.43}$, S316$^{7.46}$ (like compound 3) thereby decreasing total predicted binding by 19.2 kcal/mol compared to compound 1. Compared to compound 3, we find that compound 5 has worse interactions with L79$^{2.57}$ and V83$^{2.61}$. Although our predictions do not match the precise experimental improvement of compound 5 compared to compound 3, we can explain the difference in predicted binding as due to not re-docking the ligands to the binding site.

Overall, other than compound 4, we find a reasonable match between the predicted binding decrease and experimental binding decrease. Compound 3 and 5 have similar computational binding energies as do the experiments, while compound 2 is predicted to have greatly decreased binding to the protein, as in the experiments. Most notably, we find that:

- modifying the sulfonyl group loses favorable interactions with the receptor on TM2 and TM7, notably K76$^{2.54}$ on TM2 and S316$^{7.46}$ on TM7, leading to less favorable binding sites

- keeping the chlorine and fluorine substituents in the ligand is more optimal, since more bulky substituents at these positions cause clashes with neighboring residues on the receptor, leading to a more unfavorable binding mode.

G. Advantages of an Ensemble Based Approach for Understanding Ligand Binding

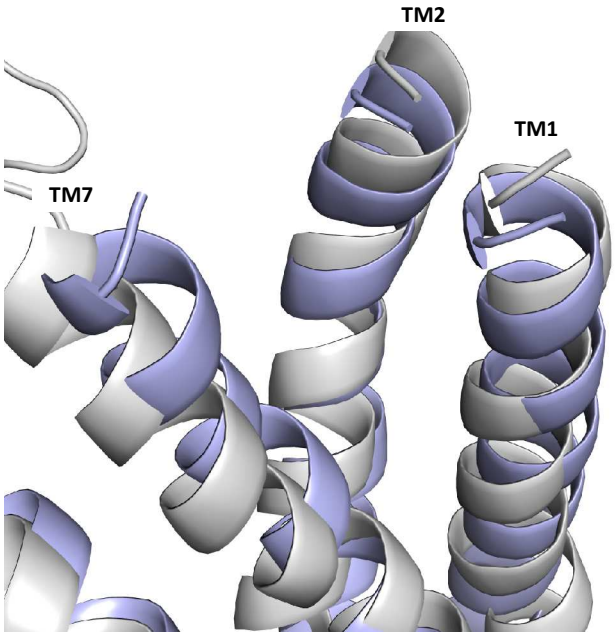


Figure 11. Superimposed binding site of 16th ranked SuperBiHelix structure based on previously predicted DP by neutral total energies (**purple**) on antagonist-receptor complex (**gray**) for selected receptor relaxed through dynamics.

Generally, GPCR binding sites have high conformational flexibility. As Latorraca et. al[54] point out, the binding pocket may be highly flexible, even within a single conformation state as seen in the agonist-bound A_{2A} receptor. Thus, docking to a single structure is limiting, since it underestimates the flexibility of the receptor binding pocket. A strength of GEnSeMBLE, compared to other methodologies including MembStruk[15], is that it provides an *ensemble* of low energy structures, allowing ligands to select their preferred conformation from the ensemble. For instance, compound 2 (Figure 11) prefers the 16th structure from the ensemble. Although

compound 2 has a highly unfavorable binding mode with the selected receptor from SuperbiHelix that has undergone dynamics (Figure 10), we find that compound 2 can bind more favorably to the 16th SuperBiHelix structure.

The binding mode with compound 2 and the selected receptor that has undergone dynamics is particularly unfavorable (Figure 10) due to the clash between the antagonist and the residues on TM1 and TM2, namely F27$^{1.43}$, S80$^{2.58}$. For the 16th ranked SuperBiHelix structure based on neutral total energies (Figure 11), we predict the antagonist would have more room to settle in the pocket, when comparing the relative positions of TM1 and TM2 in the binding pocket of the superimposed structures. Since TM7 is closer to the center of the binding pocket, we also anticipate that the R310$^{7.40}$ residue, seen to interact with the antagonist in most of our receptor-ligand complexes would pull the antagonist towards TM7, thereby minimizing the clash of the ligand with TM1 and TM2, which are more spaced apart in the binding pocket. This would allow for a more favorable binding mode that minimizes steric clashes with TM1 and TM2.

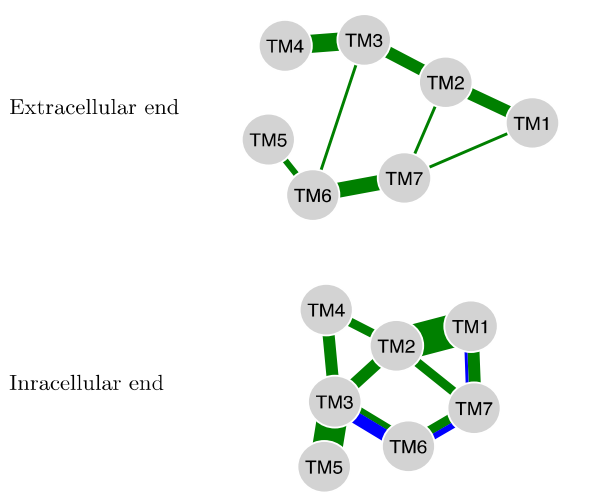
H. New Insights into hDP Structure and Activation:

We find that our current final binding mode has many of the same important interactions as DP-Li/T. Thus both have a similar stable 1-2-7 network N34$^{1.50}$- D72$^{2.50}$ - K76$^{2.54}$ - S316$^{7.46}$. We find interactions of the ligand with K2.54 and R7.40, just as in DP-Li.

An important structural improvement of our structure compared to the previous DP-Li model is in the stability of TM4 and TM5. Our DuplexBiHelix sampling found TM4 eta rotated angle by 270 degrees, thereby allowing the hydrophobic residues on TM4 to interact more freely. This rotation enabled TM4 to form other interactions with TM5. For instance, we found a stable bond formed during MD between Y199$^{5.31}$ (semi-conserved in prostaglandin receptor family) and Y203$^{5.35}$ and G172$^{4.65}$. This is consistent with Martin et. al's[25] studies on the IP receptor where it was found that Y199$^{5.31}$ comprises an aromatic cluster of residues that provide intra and inter-helical stability for TM4 and TM5, thereby maintaining the receptor in an inactive conformation. This differs from DP-Li, where Y199$^{5.31}$ faced outside the helical bundle, meaning it could interact with no other residues.

DP-Li[15] predicted a strong hydrogen bond between Q122$^{3.42}$-D319$^{7.49}$ that stabilizes the inactive conformation. In our current predictions, we find that Q122$^{3.42}$ forms alternate interactions with TM2 and TM4. Within the first 1.5 ns of MD, we find that Q122$^{3.42}$ forms interactions with S160$^{4.53}$, S157$^{4.50}$, and L64$^{2.42}$. These interactions stabilize to within 3 Å and are maintained during the 100 ns trajectory. Q122$^{3.42}$ and S157$^{4.50}$ are unique to hDP receptor and S157$^{4.50}$and L64$^{2.42}$ are semi-conserved in the prostaglandin family. Therefore, we can expect these interactions only in hDP. Since TM3[34] movement is implicated in activation for rhodopsin, β_2-adrenergic receptors, and other Class A GPCRs, we hypothesize these interactions stabilize

the inactive conformation, where activation is characterized by movement of TM3, thereby breaking bonds with TM2, TM4, and TM7.



Inter-helical Contacts in Inactive Structures	Contact Residues in Class A GPCRs	Contact Residues in hDP Receptor
3.43 - 6.40	L3.43-M/I/V6.40	L123 - M270
3.43 - 6.41	L3.43-M/L/V6.41	L123 - T271
3.46 - 6.37	L/I/M3.46 - V/I/L6.37	M126 - L267
3.50 - 6.37	R3.50 - V/I/L6.37	C130 - L267
6.36 - 7.53	M/T6.36 - Y7.53	L266 - F323
6.40 - 7.49	M/I/V6.40 - N7.49	M270 - D319
1.53 - 7.53	V1.53 - Y7.53	A37 - F323

Figure 12[35,46]. The group of interhelical interactions that switch partners upon activation in other class A receptors, namely rhodopsin, β2AR, and M2. The table on the right shows the conserved residues that characterize Class A receptors and their translation to the hDP receptor.

We also compared the contacts present in inactive structures in other Class A GPCRs (Figure 12) with contacts found in hDP receptor, which has not been characterized before in the literature. We find many variants of hDP receptor compared to the other Class A receptors.

We find that L123$^{3.43}$ is within 3-5 Å of M270$^{6.40}$ and T271$^{6.41}$ throughout the MD trajectory, allowing it to form inter-helical contacts with both residues. L3.43, M6.40, and T6.41 are all semi-conserved in the prostaglandin receptor family. Also, we find a strong hydrogen bond between L123$^{3.43}$ and S119$^{3.39}$, which is also semi-conserved in prostaglandins.

In other Class A GPCRs, activation is characterized by movement of TM3 and TM6 along with the re-arrangement of specific hydrophobic residues[36], specifically L3.43, F6.44, and

X6.40. We find L3.43 interacts strongly with S3.39, thereby anchoring S3.39 to interact with D7.49, an interaction implicated in stabilizing the receptor in the inactive conformation. We expect activation to be characterized by increasing bond distance of L3.43 and M6.40, T6.41. Since L3.43 has also been implicated in holding the receptor in the inactive state for other Class A GPCRs like CB1, β2-AR[37,38], we propose mutation experiments with L3.43 to understand hDP receptor constitutive activity.

We find M126$^{3.46}$ makes an inter-helical contact with L266$^{6.36}$ instead of L267$^{6.37}$. We find M3.46 and L6.37 are fully-conserved, while L6.36 is semi-conserved. Since M3.46 is fully conserved in the prostaglandin family, we anticipate its inter-helical contact with TM6 is characteristic of activation for all receptors in the prostaglandin family. Unlike other Class A GPCRs, we find that hDP receptor lacks the 3-6 ionic lock, with a C130$^{3.50}$ instead of R3.50. We find that C3.50 does not form an interaction with L6.37 (Figure 12). Instead, at the intracellular end, C3.50 has inter-helical contact with H263$^{6.33}$, semi-conserved in the prostaglandin family.

Another significant difference in hDP receptor is the DPxxF motif instead of the NPxxY motif on TM7, conserved on Class A GPCRs. In hDP, we find L6.36 has no contact with F7.53, unlike in other class A GPCRs where L6.36 interacts with Y7.53. Additionally, we find that M270$^{6.40}$ has no interaction with D319$^{7.49}$, which is typically N319$^{7.49}$ in other Class A GPCRs. Instead, we find M270$^{6.40}$ interacts strongly with L266$^{6.36}$ and F274$^{6.44}$, which is consistent with our understanding of the hydrophobic residue re-arrangement activation mechanism seen in other Class A GPCRs[36]. Also, we find a stable contact between F6.44 and D7.49, which has been shown to be uniquely involved in prostaglandin receptor activation[24]. Therefore, it is plausible that M6.40, L6.36, and F6.44 anchor the receptor in the inactive conformation, and their movement is characteristic of activation.

Additionally, the WxP motif in other Class A GPCRs aligns to $S^{6.48}XP$ on hDP receptor . We find that S6.48 interacts with F6.44. This interaction is significant since it has been shown in Rhodopsin, β_2-AR, $A_{2A}R$[39] that both residues are a part of a receptor "transmission switch", where each residue movement is a part of activation. We also find that F6.44 forms an inter-helical contact with D7.49. Since D7.49 is universally conserved in prostaglandin receptors and is implicated in activation[24], we expect this contact is characteristic of the inactive conformation. We observe no interaction between F6.44 and A1.53, which is plausible especially since the prostaglandin receptor family has A1.53 instead of the conserved V1.53 in other Class A receptors (Figure 12). Hence, we hypothesize that mutation experiments with F6.44 and S6.48 in hDP receptor will increase receptor constitutive activity.

V. Summary

In summary, we report here four new results.

One, we find a structure for the 3D protein-antagonist bound complex for the human DP GPCR that agrees well with experimental understanding of prostaglandin receptors and with DP-Li[15]. For example, the DP receptor has the highly conserved 1-2-7 hydrogen bond coupling between $N34^{1.50}$ – $D72^{2.50}$ – $S316^{7.46}$. Our final structure was validated against published experimental mutagenesis analysis and with structure activity relationships of various antagonists with the receptors. Our calculations found interactions between the ligand and $K76^{2.54}$, $R310^{7.40}$, $Y87^{2.65}$, $S313^{7.43}$, $S19^{1.35}$, and other neighboring residues that contribute to stabilization of the binding pocket.

Two, we report the DuplexBiHelix refinement of the SuperBiHelix procedure to ensure optimum 1-2-7 couplings for class A GPCRs. Many class A GPCRs have the conserved $Asp^{2.50}$ in the middle of TM2, with usually no nearby positive residue with which to form a salt bridge.

Thus, DuplexBiHelix, which first samples the 1-2-7 helices, can capture the important configurations that arise from the positive center interactions with $Asp^{2.50}$. We anticipate that DuplexBiHelix will be useful predicting 3D structure of other GPCRs.

Three, when the binding sites of various antagonists are selected from molecular dynamics, we find reasonable agreement in the binding energies between predictions and experiment. Our calculated binding modes show differences among antagonists in calculated binding energy based on steric clashes or loss of specific favorable interactions with the antagonist and residues on receptor. Further, this analysis highlights the strength in our ensemble-based approach compared to previous structural prediction approaches like MembStruk[15], since it allows us to analyze the preferential binding of different ligands to other low energy receptors predicted in SuperBiHelix. We have reported these structures in the Supplementary Information section.

Four, our predictions are consistent with previous suggestions for an activation mechanism[15] in that the inactive structure has strong 2-7 and 3-7 interactions between K2.54 and S7.46 and between $S119^{3.39}$ and $D319^{7.49}$, respectively. We also find in our final DP structures other interactions involving hydrophobic residues (L3.43, M6.40, and F6.44) that seem to anchor the receptor in the inactive conformation. Mutation experiments with K2.54, L3.43, M6.40, and F6.44 may help elucidate hDP receptor activity.

Overall, the data seems to suggest that the predicted 3D protein models are sufficiently accurate for further medication development of anti-inflammatory drugs for arthiritis, asthma, allergies, and cardiovascular disease.

ASSOCIATED CONTENT

Supporting Information.

The following files are available free of charge.

DP SI.docx contains Figure S1-S4, Table S1-S2

DP Structures.txt contains the following structures: Li et. al (15) antagonist-receptor complex, Antagonist-Receptor Complex from CS-HS methods, Other Possible Favorable Receptor Conformation

AUTHOR INFORMATION

Corresponding Author

William A. Goddard III (wagoddard3@gmail.com)

ACKNOWLEDGMENTS

We thank Prof. Youyong Li of Soochow University, PRC, for providing all the structure files from previous studies. We thank Cargill Corporation for providing support. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562.

REFERENCES

1. Hata AN, Breyer RM. Pharmacology and signaling of prostaglandin receptors: multiple roles in inflammation and immune modulation. Pharmacol Ther 2004;103: 147–66.

2. Pettipher, R., Hansel, T. T. & Armer, R. Antagonism of the prostaglandin D2 receptors DP1 and CRTH2 as an approach to treat allergic diseases. Nature Rev. Drug Discov. 6, 313–325 (2007).

3. Woodward DF, Jones RL, Narumiya S. International union of basic and clinical pharmacology. LXXXIII: Classification of prostanoid receptors, updating 15 years of progress. Pharmacol Rev 2011; 63:471.

4. Urade, Y. & Hayaishi, O. Prostaglandin D2 and sleep/wake regulation. Sleep Med Rev 15, 411–418 (2011).

5. Tsuboi, Kazuhito, Yukihiko Sugimoto, and Atsushi Ichikawa. "Prostanoid Receptor Subtypes." *Prostaglandins & Other Lipid Mediators* 68-69 (2002): 535-56.

6. Breyer, R. M., Bagdassarian, C. K., Myers, S. A., & Breyer, M. D. (2001). Prostanoid receptors: Subtypes and signaling. Annu Rev Pharmacol Toxicol, 41, 661–690.

7. Tan J, Abrol R, Trzaskowski B, Goddard WA 3rd (2012) 3D Structure Prediction of TAS2R38 Bitter Receptors Bound to Agonists Phenylthiocarbamide (PTC) and 6-n-Propylthiouracil (PROP). *J Chem Inf Model* 52: 1875–1885.

8. Rosenkilde, Mette M. "High Constitutive Activity of a Virus-Encoded Seven Transmembrane Receptor in the Absence of the Conserved DRY Motif (Asp-Arg-Tyr) in Transmembrane Helix 3." *Journal of Molecular Pharmocology* 68.1 (2005): 11-18. Print.

9. J. A. Ballesteros and H. Weinstein, “Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors,” in Receptor molecular biology, S. C. Sealfon, Ed., Academic Press, New York, NY, USA, 1995.

10. Abrol, R., Bray, J.K., and Goddard III, W.A. (2011) BiHelix: Towards *de novo* Structure Prediction of an Ensemble of G protein Coupled Receptor Conformations, *Proteins: Structure, Function, and Bioinformatics.*

11. Kim, S.K., Riley, L., Abrol, R., Jacobson, K. A., & Goddard, W. A., 3rd (2011). Predicted structures of agonist and antagonist bound complexes of adenosine A3 receptor. *Proteins: Structure, Function, and Bioinformatics*, 79(6), 1878–1897.

12. Kim SK, Li Y, Abrol R, Heo J, Goddard III, WA. (2011) Predicted structures and dynamics for agonists and antagonists bound to serotonin 5-HT2B and 5-HT2C receptors. *J. Chem. Inf. Model*. 51(2):420–433.

13. Kim SK, Fristrup P, Abrol R, Goddard III, W.A. (2011) Structure-based prediction of subtype selectivity of histamine H3 receptor selective antagonists in clinical trials. *J. Chem. Inf. Model.* 51(12): 3262–3274.

14. Kam, V.W.T., Goddard III, W.A., 2008. Flat-bottom strategy for improved accuracy in protein side-chain placements. J. Chem. Theor. Comput. 4, 2160–2169.

15. Li, Y., Zhu, F., Vaidehi, N., Goddard III, W.A, Sheinerman, F., Reiling, S., et al. (2007). Prediction of the 3D structure and dynamics of human DP G-protein coupled receptor bound to an agonist and an antagonist. *Journal of the American Chemical Society*, 129(35), 10720–10731.

16. Labrecque P, Roy SJ, Frechette L, Iorio-Morin C, Gallant MA, et al. (2013). Inverse agonist and pharmacochaperone properties of MK-0524 on the Prostanoid DP1 receptor. PLoS One 8.

17. Kobayashi T, Kiriyama M, Hirata T, Hirata M, Ushikubi F, Narumiya S. Identification of domains conferring ligand binding specificity to the prostanoid receptor. Studies on chimeric prostacyclin/prostaglandin D receptors. *J Biol Chem* 1997; 272:15154 –15160.

18. Stitham, Jeremiah, Aleksandar Stojanovic, Lauren A. Ross, Anthony C. Blount, and John Hwa. "Clusters of Transmembrane Residues Are Critical for Human Prostacyclin Receptor Activation." *Biochemistry* 43.28 (2004): 8974-986. Print.

19. Kedzie, K. M., Donello, J. E., Krauss, H. A., Regan, J. W. and Gil, D. W. (1998) A single amino-acid substitution in the EP2 prostaglandin receptor confers responsiveness to prostacyclin analogs. *Mol. Pharmacol*.

20. Huang, Chifu, and Hsin-Hsiung Tai. (1995) "Expression and Site-directed Mutagenesis of Mouse Prostaglandin E2 Receptor EP3 Subtype in Insect Cells." *Biochemistry Journal* 307: 493-98. Print.

21. Neuschäfer-Rube F, Engemaier E, Koch S, Boer U, Püschel GP. (2003) Identification by site-directed mutagenesis of amino acids contributing to ligand-binding specificity or signal transduction properties of the human FP prostanoid receptor. *Biochem J*. 371:443-9.

22. Funk CD, Furci L, Moran N, FitzGerald GA (1993) Point mutation in the seventh hydrophobic domain of the human thromboxane A2 receptor allows discrimination between agonist and antagonist binding sites. *Mol Pharmacol* 44:934–939.

23. Probst, W. C., L. A. Snyder, D. I. Schuster, J. Brosius, and S. C. Sealfon. Sequence alignment of the G-protein coupled receptor superfamily. *DNA Cell Biol*. 11:1–20 (1992).

24. Audoly, L. and Breyer, R. M. (1997) Substitution of charged amino acid residues in transmembrane regions 6 and 7 affect ligand binding and signal transduction of the prostaglandin EP3 receptor. *Mol. Pharmacol*. 51, 61–68.

25. Martin KA, Gleim S, Elderon L, Fetalvero K, Hwa J (2009) The human prostacyclin receptor from structure function to disease. *Prog Mol Biol Transl Sci* 89: 133–166.

26. Stitham J, Arehart EJ, Gleim SR, Douville KL, Hwa J. Human prostacyclin receptor structure and function from naturally-occurring and synthetic mutations. *Prostaglandins Other Lipid Mediat* 2007;82:95–108.

27. Stitham J, Stojanovic A, Merenick BL, O'Hara KA, Hwa J. The unique ligand-binding pocket for the human prostacyclin receptor. Site-directed mutagenesis and molecular modeling. *J Biol Chem* 2003; 278: 4250–7.

28. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 1998;102:3586–3616.

29. Liu, Wet al. Structural basis for allosteric regulation of GPCRs by sodium ions. *Science* 337, 232–236 (2012).

30. Stitham J, Arehart E, Elderon L, Gleim SR, Douville K, Kasza Z, Fetalvero K, MacKenzie T, Robb J, Martin KA, et al. (2011) Comprehensive biochemical analysis of rare prostacyclin receptor variants: study of association of signaling with coronary artery obstruction. *J Biol Chem* 286: 7060 –7069.

31. Bill, A., Rosethorne, E. M., Kent, T. C., Fawcett, L., Burchell, L., van Diepen, M. T., Marelli, A., Batalov, S., Miraglia, L., Orth, A. P., Renaud, N. A., Charlton, S. J., Gosling, M., Gaither, L. A., and Groot-Kormelink, P. J. (2014) High throughput mutagenesis for identification of residues regulating human prostacyclin (hIP) receptor expression and function. *PLoS One* 9, e97973.

32. Sturino CF, O'Neill G, Lachance N, Boyd M, Berthelette C, Labelle M, Li L, Roy B, Scheigetz J, Tsou N, Aubin Y, Bateman KP, Chauret N, Day SH, Lévesque JF, Seto C, Silva JH, Trimble LA, Carrière MC, Denis D, Greig G, Kargman S, Lamontagne S, Mathieu MC, Sawyer N, Slipetz D, Abraham WM, Jones T, McAuliffe M, Piechuta H, Nicoll-Griffith DA, Wang Z, Zamboni R, Young RN, Metters KM: Discovery of a potent and selective prostaglandin D2 receptor antagonist. *J Med Chem* 2007; 50:794– 806.

33. Sturino CF, O'Neill G, Lachance N, Boyd M, Berthelette C, Labelle M, Li L, Roy B, Schei- getz J, Tsou N, Aubin Y, Bateman KP, Chaur- et N, Day SH, Levesque J, Seto C, Silva JH, Trimble LA, Carriere M, Denis D, Greig G, Kargman S, Lamontagne S, Mathieu M, Saw-

yer N, Slipetz D, Abraham WM, Jones T, McAuliffe M, Piechuta H, Nicoll-Griffith DA, Wang Z, Zamboni R, Young RN, Metters KM. Discovery of a potent and selective prosta- glandin D2 receptor antagonist, [(3R)-4- (4-chlorobenzyl)-7-fluoro-5-(methylsulfonyl)- 1,2,3,4-tetrahydrocyclopenta[b]indol-3-yl]- acetic acid (MK-0524). *J Med Chem* 2007; 50(4):794–806.

34. Rosenbaum, D. M., Rasmussen, S. G. & Kobilka, B. K. The structure and function of G-protein-coupled receptors. *Nature* 459, 356–363 (2009).

35. Cvicek, Vaclav (2015) *Structure Prediction of G-Protein Coupled Receptors.* Dissertation (Ph.D.), California Institute of Technology. http://resolver.caltech.edu/CaltechTHESIS:02042015-031802985

36. Tehan, B. G.; Bortolato, A.; Blaney, F. E.; Weir, M. P.; Mason, J. S. Unifying Family A GPCR Theories of Activation. *Pharmacol. Ther*. 2014, 143, 51−60.

37. Tao, Y. X., Abell, A. N., Liu, X., Nakamura, K., & Segaloff, D. L. (2000). Constitutive activation of G protein-coupled receptors as a result of selective substitution of a conserved leucine residue in transmembrane helix III. *Mol Endocrinol* 14, 1272–1282.

38. D'Antona, A.M., Ahn, K. H., Wang, L., Mierke, D. F., Lucas-Lenard, J., & Kendall, D. A. (2006). A cannabinoid receptor 1 mutation proximal to the DRY motif results in constitutive activity and reveals intramolecular interactions involved in receptor activation. *Brain Res* 1108, 1–11.

39. Venkatakrishnan, A. J. et al. Molecular signatures of G-protein-coupled receptors. *Nature* 494, 185–194 (2013).

40. Wimley, W. C.; Creamer, T. P.; White, S. H. Solvation Energies of Amino Acid Side Chains and Backbone in a Family of Host-Guest Pentapeptides. *Biochemistry* 1996, 35, 5109−5124.

41. Palczewski, K. et al. Crystal structure of rhodopsin: A G-protein-coupled receptor. *Science* 289, 739–745 (2000).

42. Rasmussen, S.G. et al. (2007) Crystal structure of the human β_2 adrenergic G-protein-coupled receptor. *Nature* 450, 383–387.

43. Angel, T. E., Chance, M. R. & Palczewski, K. Conserved waters mediate structural and functional activation of family A (rhodopsin-like) G protein-coupled receptors. *Proc. Natl Acad. Sci*. USA 106, 8555–8560 (2009).

44. Yuan, S., Vogel, H. & Filipek, S. The role of water and sodium ions in the activation of the m-opioid receptor. *Angew. Chem*. 52, 10112–10115 (2013).

45. Tehan, B. G.; Bortolato, A.; Blaney, F. E.; Weir, M. P.; Mason, J. S. Unifying Family A GPCR Theories of Activation. *Pharmacol. Ther*. 2014, 143, 51−60.

46. Cvicek, V., Goddard, W. A., & Abrol, R. (2016). Structure-Based Sequence Alignment of the Transmembrane Domains of All Human GPCRs: Phylogenetic, Structural and Functional Implications. *PLoS Computational Biology* (Vol. 12). http://doi.org/10.1371/journal.pcbi.1004805.

47. Humphrey, W., Dalke, A. & Schulten, K. VMD – Visual Molecular Dynamics. *J. Mol. Graphics* **14**, 33-38 (1996).

48. Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L. & Schulten, K. Scalable molecular dynamics with NAMD. *J. Comp. Chem.* **26**, 1781-1802 (2005).

49. MacKerell, Jr., A. D., Bashford, D., Bellott, M., Dunbrack Jr., R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher, III, W.E., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D. & Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586-3616 (1998).

50. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. & Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79,** 926–935 (1983).

51. Darden, T.A., York, D.M., Pederson, L.G. Particle mesh ewald – An N.Log(N) method for ewald sums in large systems. *J. Chem. Phys.* 98, 10089 (1993).

52. Quigley, D. & Probert, M.I.J. Langevin dynamics in constant pressure extended systems. *J. Chem. Phys.* 120, 11432 (2004).

53. Goddard III, W. A., Kim, S., Li, Y., Trzaskowski, B., & Griffith, A. R. (2010). Predicted 3D structures for adenosine receptors bound to ligands : Comparison to the crystal structure. *Journal of Structural Biology*, *170*(1), 10–20. http://doi.org/10.1016/j.jsb.2010.01.001.

54. Latorraca, N. R., Venkatakrishnan, A. J., & Dror, R. O. (2016). GPCR Dynamics: Structures in Motion. *ACS Chemical Reviews*. http://doi.org/10.1021/acs.chemrev.6b00177.

55. Abrol, R., Griffith, A. R., Bray, J. K., & Goddard III, W.A. (2012). Structure prediction of G protein-coupled receptors and their ensemble of functionally important conformations. In N. Vaidehi & J. Klein-Seetharaman (Eds.), *Membrane protein structure: Methods and protocols*, Vol. 914, (pp. 237–254). New York: Humana.

56. Bray, JK and Goddard III, WA; The structure of human serotonin 2c G-protein-coupled receptor bound to agonists and antagonists, *J. Mol. Graph. Model*. 27 (1): 66-81 (2008)

57. Griffith, Adam Reid (2017) *DarwinDock and GAG-Dock: Methods and Applications for Small Molecule Docking.* Dissertation (Ph.D.), California Institute of Technology. doi:10.7907/Z91Z42GS.http://resolver.caltech.edu/CaltechTHESIS:06122017-230026717

58. Surgand, J., Rodrigo, J., Kellenberger, E., & Rognan, D. (2006). A Chemogenomic Analysis of the Transmembrane Binding Cavity of Human G-Protein-Coupled Receptors. *PROTEINS: Structure, Function, and Bioinformatics*, *538*(April 2005), 509–538. http://doi.org/10.1002/prot.20768.

59. Kobayashi, Takuya; Ushikubi, Fumitaka; Narumiya, S. (2000). Amino Acid Residues Conferring Ligand Binding Properties of Prostaglandin I and Prostaglandin D Receptors. *Journal of Biological Chemistry*, *275*(32), 24294–24303. http://doi.org/10.1074/jbc.M002437200.

60. Zuercher, W. J., Elkins, J. M., & Knapp, S. (2016). The Intersection of Structural and Chemical Biology - An Essential Synergy. *Cell Chemical Biology*, *23*(1), 173–182. http://doi.org/10.1016/j.chembiol.2015.12.005.

61. Abrol R, et al. (2014) Ligand- and mutation-induced conformational selection in the CCR5 chemokine G protein-coupled receptor. *Proceedings of the National Academy of Sciences of the United States of America* 111(36):13040–13045.

62. Scott CE, Abrol R, Ahn KH, Kendall DA, Goddard WA, III (2013) Molecular basis for dramatic changes in cannabinoid CB1 G protein-coupled receptor activation upon single and double point mutations. *Protein Sci* 22(1):101–113.

63. Li, Q., Kim, S.-K., Goddard, W. A., Chen, G. & Tan, H. (2015) Predicted structures for kappa opioid g-protein coupled receptor bound to selective agonists. *J. Chem. Inf. Model*. 55, 614–627.

64. A. Kirkpatrick, J. Heo, R. Abrol, and W. A. Goddard III, (2012) "Predicted structure of agonist-bound glucagon-like peptide 1 receptor, a class B G protein-coupled receptor,"

Proceedings of the National Academy of Sciences of the United States of America, vol. 109, no. 49, pp. 19988–19993.

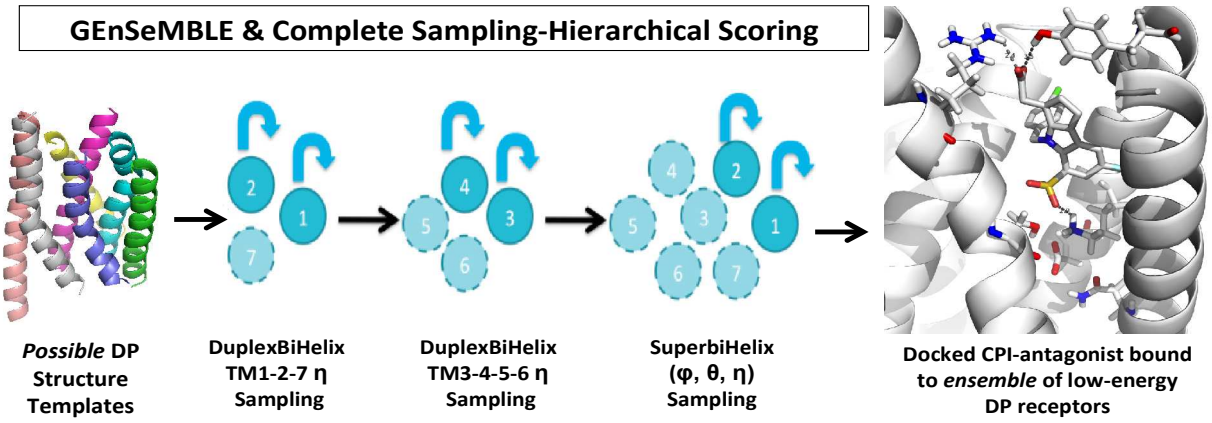
65. Kim, S., Chen, Y., Abrol, R., Goddard, W. A., & Guthrie, B. (2017). Activation mechanism of the G protein-coupled sweet receptor heterodimer with sweeteners and allosteric agonists. *Proceedings of the National Academy of Sciences*, *114*(10), 2568–2573. http://doi.org/10.1073/pnas.1700001114.

66. Bray, J. K., Abrol, R., Iii, W. A. G., Trzaskowski, B., & Scott, C. E. (2013). SuperBiHelix method for predicting the pleiotropic ensemble of G-protein – coupled receptor conformations. *Proceedings of the National Academy of Sciences*, 72–78. http://doi.org/10.1073/pnas.1321233111

for Table of Contents use only

The Predicted 3D Structure of Human DP Prostaglandin G Protein-Coupled Receptor Bound to CPI Antagonist, predicted using the DuplexBiHelix modification of the GEnSeMBLE method

AUTHOR NAMES: Vishnu Shankar, William A. Goddard III[1], Soo-Kyung Kim, Ravinder Abrol, Fan Liu*



ABSTRACT: Prostaglandins play a critical physiological role in both cardiovascular and immune systems, acting through its interactions with 9 prostanoid G protein-coupled receptors (GPCRs). These receptors are important therapeutic targets for a variety of diseases including arthritis, allergies, type 2 diabetes, and cancer. The DP prostaglandin receptor is of interest because it has unique structural and physiological properties. Most notably, DP does not have the 3-6 ionic lock common to Class A GPCRs. However, the lack of x-ray structures for any of the 9 prostaglandin GPCRs hampers the application of structure-based drug design methods to

develop more selective and active medications to specific receptors. We predict here 3D structures for the DP prostaglandin GPCR, based on the GEnSeMBLE complete sampling with hierarchical scoring (CS-HS) methodology. This involves evaluating the energy of 13 trillion packings to finally select the best 20 that are stable enough to be relevant for binding to antagonists, agonists, and modulators. To validate the predicted structures, we predict the binding site for the Merck cyclopentanoindole (CPI) selective antagonist docked to DP. We find that the CPI binds vertically in the 1-2-7 binding pocket, interacting favorably with residues R310$^{7.40}$ and K76$^{2.54}$ with additional interactions with S313$^{7.43}$, S316$^{7.46}$, S19$^{1.35}$, etc. This binding site differs significantly from that of antagonists to known Class A GPCRs where the ligand binds in the 3-4-5-6 region. We find that the predicted binding site leads to reasonable agreement with experimental Structure Activity Relationships (SAR). We suggest additional mutation experiments including K76$^{2.54}$, E129$^{3.49}$, L123$^{3.43}$, M270$^{6.40}$, F274$^{6.44}$ to further validate the structure, function, and activation mechanism of receptors in the prostaglandin family. Our structures and binding sites are largely consistent and improve upon the predictions by Li et. al[15] that used our earlier MembStruk prediction methodology.